(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2001 (20.12.2001)

PCT

(10) International Publication Number
WO 01/96547 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US01/19444

(22) International Filing Date: 14 June 2001 (14.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/212,073	15 June 2000 (15.06.2000)	US
60/213,467	23 June 2000 (23.06.2000)	US
60/215,651	30 June 2000 (30.06.2000)	US
60/216,605	7 July 2000 (07.07.2000)	US
60/218,372	13 July 2000 (13.07.2000)	US
60/228,056	25 August 2000 (25.08.2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **GREENWALD, Sara, R.** [US/US]; 21 Bucareli Drive, San Francisco, CA 94132 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **KEARNEY, Liam** [IE/US]; 50 Woodside Avenue, San Jose, CA 94127 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **HE, Ann** [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062 (US). **HAFALIA, April** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). **LO, Terence, P.** [CA/US]; 1451 Beach Park Blvd., Apt. 115, Foster City, CA 94404 (US). **KHAN, Farrah** [IN/US]; 3617 Central Road #102, Glenview, IL 60025 (US). **RECIPON, Shirley, A.** [US/US]; 85 Fortuna Avenue, San Francisco, CA 95115 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **POLICKY, Jennifer, L.** [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **GRETHER, Megan** [US/US]; 66 Nordhoff Street, San Francisco, CA 94131 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue, #23, Mountain View, CA 94043 (US). **BATRA, Sajeev** [US/US]; 555 El Camino Real, #709, San Leandro, CA 94577 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) **Title:** HUMAN KINASES

(57) **Abstract:** The invention provides human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating or prevention disorders associated with aberrant expression of PKIN.

WO 01/96547 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

HUMAN KINASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and

sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

Protein Tyrosine Kinases

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased

tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on

the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α, β, γ, δ, and ε. Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., supra).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the Drosophila circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro,* CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomian transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle

contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and seratonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.

Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the

dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakarocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from Drosophila polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of

a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon-γ induced apoptosis (Sanjo et al., supra). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., supra). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., supra).

Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell. Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP_2). PIP_2 is then cleaved into inositol triphosphate (IP_3) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism.

PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP_2) to PI (3,4,5) P_3 (PIP_3). PIP_3 then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, supra).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring

high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as $p21^{ras}$ known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of $p21^{ras}$ and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for de novo synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

The discovery of new human kinases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of

nucleic acid and amino acid sequences of human kinases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," "PKIN-11," "PKIN-12," "PKIN-13," "PKIN-14," "PKIN-15," "PKIN-16," "PKIN-17," "PKIN-18," "PKIN-19," "PKIN-20," "PKIN-21," "PKIN-22," "PKIN-23," "PKIN-24," "PKIN-25," and "PKIN-26." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-26. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:27-52.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a

transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to

said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a

composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID

NO:1-26. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:27-52, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide

sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is

not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PKIN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PKIN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic

molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or

translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His

Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For

example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:27-52 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:27-52, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:27-52 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:27-52 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:27-52 and the region of SEQ ID NO:27-52 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-26 is encoded by a fragment of SEQ ID NO:27-52. A fragment of SEQ ID NO:1-26 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-26. For example, a fragment of SEQ ID NO:1-26 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-26. The precise length of a fragment of SEQ ID NO:1-26 and the region of SEQ ID NO:1-26 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in

a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

"Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU

primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,

chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection,

electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human human kinases (PKIN), the polynucleotides encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI).

Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are human kinases. For example, SEQ ID NO:4 is 94% identical to rat serine/threonine kinase (GenBank ID g2052189) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a protein kinase. In an alternate example, SEQ ID NO: 23 is 88% identical to murine protein kinase (GenBank ID g406058) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:23 also contains an eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:23 is a protein kinase. In an alternate example, SEQ ID NO:6 is 85% identical to rabbit myosin light chain kinase (GenBank ID g165506) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.5e-272, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a eukaryotic protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:6 is a myosin light chain kinase. In an alternate example, SEQ ID NO:1 is 64% identical to murine serine/threonine kinase (GenBank ID g404634) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.5e-60, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a protein kinase, notably a serine/threonine kinase. In an alternate example, SEQ ID NO:19 is 49% identical to

human G-protein-coupled receptor kinase GRK4-beta (GenBank ID g992672) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.3e-129, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:19 also contains a regulator of G-protein signaling domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:19 is a G-protein-coupled receptor kinase. SEQ ID NO:2-3, SEQ ID NO:5, SEQ ID NO:7-18, SEQ ID NO:20-22 and SEQ ID NO:24-26 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-26 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:27-52 or that distinguish between SEQ ID NO:27-52 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6829315H1 is the identification number of an Incyte cDNA sequence, and SINTNOR01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 55057226H1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2954208) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including

the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is the identification number of a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	**Type of analysis and/or examples of programs**
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to

assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-52, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:27-52, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:27-52 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:27-52. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons

are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:27-52 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.)

Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of

the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PKIN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or combined with sequences

from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or

animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PKIN. Transcription of sequences encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311).

Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980)

Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of

an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which direct secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize

these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PKIN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PKIN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential

pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with lipid disorders, pancreatic islet cells, liver disease, leukocytes, umbilical endothelial cells, cancer, as well as, normal and diseased brain, renal, reproductive, bladder tumor, posterior hippocampus, kidney, small intestine, colon, and digestive tissues. Therefore, PKIN appears to play a role in cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to increase the expression or activity of PKIN.

Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental

disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder, such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine

deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such disorders include, but are not limited to, those cancer, immune disorders, disorders affecting growth and development, cardiovascular diseases, and lipid disorders described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various

disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PKIN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single

chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial

hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may

require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages

within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PKIN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a

human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PKIN may be used for the diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the

probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:27-52 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease

(MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder, such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency,

hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM_2 gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the

patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are

analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference

in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in <u>DNA Microarrays: A Practical Approach</u>, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent <u>in situ</u> hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, <u>supra</u>, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map

and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

In additional embodiments, the nucleotide sequences which encode PKIN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/212,073, U.S. Ser. No. 60/213,467, U.S. Ser. No. 60/215,651, U.S. Ser. No. 60/216,605, U.S. Ser. No. 60/218,372, and U.S. Ser. No. 60/228,056 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:27-52. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative human kinases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode human kinases, the encoded polypeptides were analyzed by querying against PFAM models for human kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as human kinases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

<u>"Stitched" Sequences</u>

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence.

Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PKIN Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:27-52 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:27-52 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:27 was mapped to chromosome 19 and SEQ ID NO:35 was mapped to chromosome 15 within the interval from 72.30 to 77.30 centiMorgans. SEQ ID NO:48 was mapped to chromosome 10 within the interval from 93.80 to 96.90 centiMorgans. SEQ ID NO:49 was mapped to chromosome 13 within the interval from 11.60 to 22.80 centiMorgans, to chromosome 17 within the interval from 0.60 to 14.80 centiMorgans, and to chromosome 20 within the interval from 57.70 to 64.10 centiMorgans. More than one map location is reported for SEQ ID NO:49, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PKIN Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was

synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:27-52 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, <u>supra</u>.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned

technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is

then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

XII. Expression of PKIN

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-

His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII where applicable.

XIII. Functional Assays

PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PKIN Specific Antibodies

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PKIN Using Specific Antibodies

Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

XVI. Identification of Molecules Which Interact with PKIN

PKIN, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PKIN Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by PKIN in the presence of gamma-labeled ^{32}P-ATP. PKIN is incubated with the protein substrate, ^{32}P-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P-ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of PKIN. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P-ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P-peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P-ATP. The reservoir of the centrifuged unit containing the ^{32}P-peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated in vitro in an assay containing PKIN, 50μl of kinase buffer, 1μg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μg ATP, and 0.5μCi [γ-^{33}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ-^{33}P]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation

counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of ^{32}P from gamma-labeled ^{32}P -ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ^{32}P-labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

XVIII. Enhancement/Inhibition of Protein Kinase Activity

Agonists or antagonists of PKIN activation or inhibition may be tested using assays described in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in PKIN activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2011384	1	2011384CD1	27	2011384CB1
2004888	2	2004888CD1	28	2004888CB1
2258952	3	2258952CD1	29	2258952CB1
7473244	4	7473244CD1	30	7473244CB1
1242491	5	1242491CD1	31	1242491CB1
2634875	6	2634875CD1	32	2634875CB1
3951059	7	3951059CD1	33	3951059CB1
7395890	8	7395890CD1	34	7395890CB1
7475546	9	7475546CD1	35	7475546CB1
7477076	10	7477076CD1	36	7477076CB1
1874092	11	1874092CD1	37	1874092CB1
4841542	12	4841542CD1	38	4841542CB1
7472695	13	7472695CD1	39	7472695CB1
7477966	14	7477966CD1	40	7477966CB1
7163416	15	7163416CD1	41	7163416CB1
7472822	16	7472822CD1	42	7472822CB1
7477486	17	7477486CD1	43	7477486CB1
3773709	18	3773709CD1	44	3773709CB1
7477204	19	7477204CD1	45	7477204CB1
3016969	20	3016969CD1	46	3016969CB1
063497	21	063497CD1	47	063497CB1
1625436	22	1625436CD1	48	1625436CB1
3330646	23	3330646CD1	49	3330646CB1
3562763	24	3562763CD1	50	3562763CB1
621293	25	621293CD1	51	621293CB1
7480774	26	7480774CD1	52	7480774CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	2011384CD1	g404634	4.50E-60	[Mus musculus] serine/threonine kinase (Bielke,W. et al (1994) Gene 139 (2), 235-239)
		g13540326	1.00E-159	[fl][Homo sapiens] serine/threonine kinase FKSG82
2	2004888CD1	g2983205	2.70E-08	[Aquifex aeolicus] ser/thr protein kinase (Deckert,G. et al (1998) Nature 392 (6674), 353-358)
		g13603881	0	[fl][Homo sapiens] serine/threonine kinase 31 (Wang,P.J. et al, (2001) Nat. Genet. 27 (4), 422-426)
3	2258952CD1	g3766209	0	[Mus musculus] IRE1 (Wang,X.Z. et al (1998) EMBO J. 17 (19), 5708-5717)
4	7473244CD1	g2052189	0	[Rattus norvegicus] serine/threonine kinase
5	1242491CD1	g2253010	8.40E-25	[Arabidopsis thaliana] MAP3K delta-1 protein kinase (Jouannic,S. et al (1999) Gene 229 (1-2), 171-181)
6	2634875CD1	g13194657	0	[fl][Homo sapiens] skeletal myosin light chain kinase
		g165506	1.50E-272	[Oryctolagus cuniculus] myosin light chain kinase (EC 2.7.1.-) (Herring,B.P. et al (1990) J. Biol. Chem. 265, 1724-1730)
7	3951059CD1	g3599507	5.00E-235	[Mus musculus] rho/rac-interacting citron kinase short isoform (Di Cunto,F. et al (1998) J. Biol. Chem. 273 (45), 29706-29711)
8	7395890CD1	g5815139	0	[Mus musculus] nuclear body associated kinase 1a
9	7475546CD1	g3435114	1.80E-50	[Homo sapiens] serine/threonine kinase ULK1 (Kuroyanagi,H. et al (1998) Genomics 51 (1), 76-85)
10	7477076CD1	g854733	6.20E-198	[Rattus norvegicus] casein kinase 1 gamma 1 isoform
11	1874092CD1	g2511715	4.00E-25	[Arabidopsis thaliana] putative phosphatidylinositol-4-phosphate
12	4841542CD1	g927732	3.30E-67	[Saccharomyces cerevisiae] Snf1p: serine/threonine protein kinase;
13	7472695CD1	g1498250	1.10E-53	[Dictyostelium discoideum] myosin light chain kinase (Tan,J.L. et al (1991) J. Biol. Chem. 266, 16044-16049)
		g12830367	0	[fl][Homo sapiens] serine/threonine kinase 33
14	7477966CD1	g3766209	0	[Mus musculus] IRE1 (Wang,X.Z. et al (1998) EMBO J. 17 (19), 5708-5717)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
15	7163416CD1	g7649810	2.10E-135	[Homo sapiens] protein kinase PAK5
		g11691855	0	[fl][Homo sapiens] pak5 protein
16	7472822CD1	g5081459	3.70E-241	[Danio rerio] p55-related MAGUK protein DLG3
17	7477486CD1	g3217028	0	[5' incom][Homo sapiens] putative serine/threonine protein kinase (Stanchi,F. et al (2001) Yeast 18 (1), 69-80)
18	3773709CD1	g3986088	6.70E-78	[Pyrococcus kodakaraensis] Glycerol Kinase
19	7477204CD1	g992672	7.30E-129	[Homo sapiens] G protein-coupled receptor kinase GRK4-beta (Premont,R.T. et al (1996) J. Biol. Chem. 271 (11), 6403-6410)
		g4001826	0	[fl][Spermophilus tridecemlineatus] G protein-coupled receptor kinase GRK7 (Weiss,E.R. et al (1998) Mol. Vis. 4, 27)
20	3016969CD1	g4521278	4.70E-45	[Homo sapiens] Trad (Kawai,T. et al (1999) Gene 227 (2), 249-255)
21	63497CD1	g1213224	0	[Rattus norvegicus] SNF1-related kinase (Becker,W. et al (1996) Eur. J. Biochem. 235 (3), 736-743)
22	1625436CD1	g4096108	1.10E-252	[Homo sapiens] proline rich calmodulin-dependent protein kinase
		g206152	0	[fl][Rattus norvegicus] calmodulin-dependent protein kinase II gamma subunit (EC 2.7.1.37) (Tobimatsu,T. et al (1988) J. Biol. Chem. 263, 16082-16086)
23	3330646CD1	g406058	0	[Mus musculus] protein kinase (Walden,P.D. et al (1993) Mol. Cell. Biol. 13, 7625-7635)
24	3562763CD1	g12830335	0	[5' incom][Homo sapiens] bA55008.2 (novel protein kinase)
		g1510182	9.80E-18	[Mus musculus] cyclin-dependent kinase 5 (Ishizuka,T. et al (1995) Gene 166 (2), 267-271)
25	621293CD1	g2649941	4.50E-23	[Archaeoglobus fulgidus] adenylate kinase (adk) (Klenk,H.P. et al (1997) Nature 390 (6658), 364-370)
26	7480774CD1	g2463542	0	"[Homo sapiens] inositol 1,4,5-trisphosphate 3-kinase"

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2011384CD1	273	Y12 Y23 T17 S144 T30 S31 S237 S253		PROTEIN KINASE DOMAIN DM00004|P27448|58-297: R16-R255	BLAST_DOMO
					Eukaryotic protein kinase domain pkinase: Y12-L267	HMMER_PFAM
					Protein kinases signatures and profile, protein_kinase_tyr.prf: Q111-G163	PROFILESCAN
					Protein Kinase ATP binding site: I18-K41	MOTIFS
					Protein Kinase (serine/threonine): L131-L143	MOTIFS
					Tyrosine kinase catalytic domain signature: PR00109:Y125-L143 Y193-S215	BLIMPS_PRINTS
2	2004888CD1	329	S190 S50 S51 T141 Y302		Eukaryotic protein kinase domain pkinase: P135-L228	HMMER_PFAM
					DM00004|P54744|13-263 PROTEIN KINASE DOMAIN: P113-L228 (P=1.1e-06)	BLAST_DOMO
3	2258952CD1	938	S207 S299 S500 S503 S580 S609 S65 S714 S814 S852 S857 T116 T128 T147 T175 T188 T202 T345 T55 T592 T658 T84 T895 T905 T936 Y146	N200	PROTEIN KINASE DOMAIN DM00004|Q09499|536-784:P534-A784	BLAST_DOMO
					KINASE; THREONINE; ATP; SERINE; DM06305|Q09499|786-924:V787-Y922	BLAST_DOMO
					PROTEIN KINASE/ENDORIBONULCEASE PUTATIVE SERINE/THREONINE PROTEIN KINASE C41C4.4 CHROMOSOME II PRECURSOR TRANSFERASE PD152704:T170-L395,L61-E163	BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					SERINE/THREONINE PROTEIN KINASE PRECURSOR TRANSMEMBRANE SIGNAL TRANSFERASE ATP-BINDING PROTEIN IRE1 GLYCOPROTEIN PD032590:W794-Y922	BLAST_PRODOM
					Tyrosine kinase catalytic domain PR00109: H639-I657, G694-L704, V716-D738	BLIMPS_PRINTS
					Protein kinases signatures and profile protein_kinase_tyr.prf: E625-G682	PROFILESCAN
					Eukaryotic protein kinase domain pkinase: F532-F793	HMMER_PFAM
					Protein_Kinase serine/theronine: I645-I657	MOTIFS
4	7473244CD1	795	S140 S2 S301 S35 S423 S468 S485 S486 S49 S524 S546 S609 S666 S671 S699 S705 S710 S776 T128 T19 T282 T324 T333 T437 T504 T511 T568 T581 T648 T657 T676 T680 T82 T9	N17 N331 N397 N398	Protein kinases signatures and profile protein_kinase_tyr.prf: Y133-G210	PROFILESCAN
					Eukaryotic protein kinase domain pkinase: Y60-M311	HMMER_PFAM
					PROTEIN KINASE DOMAIN DM00004|P27448|58-297:L62-L302	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4					KINASE SERINE/THREONINEP ROTEIN PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PUTATIVE KIN1 EMK PAR1 PD004300:G682-L795	BLAST_PRODOM
					SERINE/THREONINE KINASE PD119193:I594-P665	BLAST_PRODOM
					KINASE SERINE/THREONINE PROTEIN SERINE/THREONINE PUTATIVE TRANSFERASE ATP-BINDING PROTEIN EMK P78 CDC25C PD008571:S412-S595	BLAST_PRODOM
					KINASE SERINE/THREONINE PROTEIN PUTATIVE SERINE/THREONINE TRANSFERASE ATP-BINDING PROTEIN PAR1 KP78 EMK PD005838:M311-R411	BLAST_PRODOM
					Tyrosine kinase catalytic domain PR00109: M136-V149, Y172-L190, V238-Q260	BLIMPS_PRINTS
					Protein_Kinase_ATP binding site: I66-K89	MOTIFS
					Protein_Kinase serine/theronine: I178-L190	MOTIFS
5	1242491CD1	656	S309 S42 S540 S569 S583 S594 S654 T270 T303 T319 T366 T408 T439 T509 T526 T570 T609 T612 T623 T653	N293 N424 N437	Eukaryotic protein kinase domain: L14-V257	HMMER_PFAM
					Protein kinases signatures and profile: L99-Q151	PROFILESCAN
					Protein kinases ATP-binding region signature: L14-K35	MOTIFS
					Serine/Threonine protein kinases active-site signature: I119-L131	MOTIFS

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>5</td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:M76-Q89, Y113-L131, A183-G205, P232-S254</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P42679|236-470:L14-P252
DM00004|I49621|195-428:L14-P252
DM00004|I38044|100-349:L13-P252
DM00004|Q05609|553-797:L14-T197, L14-T253</td><td>BLAST_DOMO</td></tr>
<tr><td>6</td><td>2634875CD1</td><td>596</td><td>S107 S143 S157 S159 S184 S203 S235 S397 S460 S586 S59 T17 T224 T247 T301 T320 T351 T379 T49 Y376</td><td>N278 N416</td><td>Eukaryotic protein kinase domain:
M285-L540</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:M359-V372, F396-C414, T463-D485</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
L291-K314</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
V402-C414</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>KINASE MYOSIN LIGHT CHAIN SKELETAL MUSCLE MLCK TRANSFERASE SERINE/THREONINE CALMODULIN BINDING
PD036174:A95-M285
PD027051:L540-V596
PD029157:A2-R82, A2-S90</td><td>BLAST_PRODOM</td></tr>
</table>

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>6</td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P07313|298-541:S287-A531
DM00004|JN0583|727-969:K288-N530
DM00004|S07571|5152-5396:E289-M529
DM00004|P53355|15-257:E289-M529</td><td>BLAST_DOMO</td></tr>
<tr><td>7</td><td>3951059CD1</td><td>497</td><td>S140 S248 S308
S361 S381 S386
S410 S436 S445
S490 S81 S93
T279 T378 T83</td><td></td><td>Eukaryotic protein kinase domain:
F97-F360</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinase C terminal domain:
S361-E390</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:M174-N187, S211-V229</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
V103-K126</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
Y217-V229</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>RHO/RACINTERACTING CITRON KINASE SHORT ISOFORM
PD154232:S422-V468
PD154360:M1-M43</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>KINASE RHO ASSOCIATED COILED COIL PROTEIN FORMING RHO/RAC INTERACTING CITRON ALPHA
PD007970:Q32-D96</td><td>BLAST_PRODOM</td></tr>
</table>

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>7</td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|Q09013|83-336: V99-L349
DM00004|S42867|75-498: S101-G241, I258-S445
DM00004|S42864|41-325: E98-G241, N249-L349, D96-T153
DM00004|P38679|238-527:L102-G241, I258-L349, E86-A124</td><td>BLAST_DOMO</td></tr>
<tr><td rowspan="7">8</td><td rowspan="7">7395890CD1</td><td rowspan="7">1171</td><td rowspan="7">S121 S135 S178 S180 S254 S27 S37 S405 S649 S773 S774 S783 S788 S804 S865 S970 T119 T172 T221 T431 T450 T483 T517 T839 T867 T893 T995 T1022 S1027 S1099 Y443 Y468</td><td rowspan="7">N140 N157 N271 N480 N562 N579 N786 N963 N978 N1012</td><td>Eukaryotic protein kinase domain:
Y199-P420, R498-V527</td><td>HMMER_PFAM</td></tr>
<tr><td>Tyrosine kinase catalytic domain signature
PR00109:K314-L332</td><td>BLIMPS_PRINTS</td></tr>
<tr><td>Protein kinases ATP-binding region signature:
L205-K228</td><td>MOTIFS</td></tr>
<tr><td>Serine/Threonine protein kinases active-site signature:
L320-L332</td><td>MOTIFS</td></tr>
<tr><td>PROTEIN KINASE NUCLEAR HOMEO DOMAIN INTERACTING DNA-BINDING SERINE/THREONINE
PD141983:A573-C933
PD150874:A993-I1171</td><td>BLAST_PRODOM</td></tr>
<tr><td>PROTEIN KINASE NUCLEAR SERINE/THREONINE HOMEO DOMAIN INTERACTING DNA-BINDING SERINE/THREONINE F20B6.8
PD042899:L425-P574</td><td>BLAST_PRODOM</td></tr>
<tr><td>HOMEO DOMAIN INTERACTING PROTEIN KINASE 2 DNA-BINDING NUCLEAR PROTEIN
PD184491:E872-P961</td><td>BLAST_PRODOM</td></tr>
</table>

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>8</td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P14680|371-694: V201-P518
DM00004|Q09815|519-804: E200-L473, F500-T517
DM00004|P49657|101-409: L205-P518
DM00004|Q09690|700-985: E200-P444, F500-P518</td><td>BLAST_DOMO</td></tr>
<tr><td>9</td><td>7475546CD1</td><td>470</td><td>S134 S146 S165 S217 S219 S227 S256 S260 S339 S361 S406 S447 S462 T105 T17 T37 T61</td><td>N132</td><td>Eukaryotic protein kinase domain:
F14-V270</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:M91-H104, F127-L145, L239-F261</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases signatures and profile:
V113-P166</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
L20-K44</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
I133-L145</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>KINASE PROTEIN TRANSFERASE ATP BINDING SERINE/THREONINE RECEPTOR TYROSINE PRECURSOR TRANSMEMBRANE
PD000001:S176-P255, I15-F93, P237-W269, F117-M164, L20-K34</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P53104|26-315: P151-F261, E18-E111, F117-S147
DM00004|S54788|154-400:L20-S260
DM00004|P27448|58-297: L16-R258
DM00004|P49673|31-267: L20-I259</td><td>BLAST_DOMO</td></tr>
</table>

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>10</td><td>7477076CD1</td><td>422</td><td>S124 S150 S229 S96 T137 T14 T199 T214 T258 T269 T273 T355 T374 T417</td><td></td><td>Eukaryotic protein kinase domain pkinase:
F44-E276</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases signatures and profile:
T140-P197</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
I50-K73</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
L160-I172</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>CASEIN KINASE I GAMMA ISOFORM TRANSFERASE SERINE/THREONINE ATP BINDING MULTIGENE
PD015080:F315-T393</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>CASEIN KINASE I, GAMMA 1 ISOFORM EC 2.7.1. CKI GAMMA TRANSFERASE SERINE/THREONINE PROTEIN ATP BINDING MULTIGENE
PD049080:M1-N43</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|A56711|46-303:V46-Y304
DM00004|C56711|45-301:V46-Y304
DM00004|B56711|48-303:V46-Y304
DM00004|D56406|31-276:V46-V293</td><td>BLAST_DOMO</td></tr>
<tr><td>11</td><td>1874092CD1</td><td>240</td><td>S121 S132 S78 T197 T84</td><td></td><td>PROTEIN PHOSPHATIDYL INOSITOL 4-PHOSPHATE 5-KINASE PUTATIVE T22C1.7 ISOLOG ATPIP5K1 T4C15.16
PD149995: L13-D204</td><td>BLAST_PRODOM</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	4841542CD1	594	S108 S114 S293 S297 S305 S306 S339 S343 S382 S40 S427 S48 S489 S572 S88 S99 T193 T255 T259 T357 T477 T544 T582 Y425	N542 N87	KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: K3-S163, S178-F216, P236-W268 (P=1.2e-09)	BLAST_PRODOM
					PROTEIN KINASE DOMAIN DM00004 | P27448 | 58-297: L22-L260 DM00004 | P06782 | 57-296: L22-L260 DM00004 | JC1446 | 20-261: T24-L260 DM00004 | P54645 | 17-258: E23-L260	BLAST_DOMO
					Tyrosine kinase catalytic domain signature PR00109: M95-S108, Y131-L149, V197-H219	BLIMPS_PRINTS
					Eukaryotic protein kinase domain pkinase: Y19-V269	HMMER_PFAM
					Protein_Kinase_ATP L25-K47	MOTIFS
					Protein_Kinase_Serine/Threonine V137-L149	MOTIFS
13	7472695CD1	473	S128 S170 S208 S233 S255 S285 S30 S308 S347 S366 S379 S39 S400 S432 S46 T143 T29 T330 T371 T399 T409 T418 T469 T93	N172 N370 N397 N54	Eukaryotic protein kinase domain pkinase: Y75-L340	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13					PROTEIN KINASE DOMAIN DM00004 | S57347 | 21-266: F77-T330 DM00004 | S46283 | 13-259: G78-A331 DM00004 | S54788 | 154- 400: G78-A331 DM00004 | P28583 | 35-282: G78-A331	BLAST_DOMO
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: D197-L299, R79-D156	BLAST_PRODOM
					Tyrosine kinase catalytic domain PR00109: M151-D164, Y187-V205, C263-S285, T143-R165	BLIMPS_PRINTS
					Phosphorylase kinase family PR101049: D164-I184	BLIMPS_PRINTS
					Protein_Kinase_ATP L81-K104	MOTIFS
					Protein_Kinase_Serine/Threonine I193-V205	MOTIFS
					protein_kinase_tyrosine.profile: E173-A228	PROFILESCAN
14	7477966CD1	947	S207 S299 S508 S511 S589 S618 S65 S723 S823 S861 S866 T116 T128 T147 T175 T188 T202 T345 T55 T601 T667 T84 T904 T914 T945 Y146	N200	Eukaryotic protein kinase domain pkinase: F541-F802	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14					PROTEIN KINASE DOMAIN DM00004|Q09499|536-784: P543-A793 DM00004|P32361|676-970: V546-Q714, T722-A793	BLAST_DOMO
					do KINASE; THREONINE; ATP; SERINE; DM06305|Q09499|786-924: V796-Y931 DM06305|P32361|972-1114: Q795-L928	BLAST_DOMO
					PROTEIN KINASE/ENDORIBONULCEASE PUTATIVE SERINE/THREONINE PROTEIN KINASE C41C4.4 CHROMOSOME II PRECURSOR TRANSFERASE PD152704: T170-L395, L61-E163	BLAST_PRODOM
					SERINE/THREONINE PROTEIN KINASE PRECURSOR TRANSMEMBRANE SIGNAL TRANSFERASE ATP-BINDING PROTEIN IRE1 GLYCOPROTEIN PD032590: W803-Y931	BLAST_PRODOM
					Tyrosine kinase catalytic domain signature PR00109: H648-I666, G703-L713, V725-D747	BLIMPS_PRINTS
					Phosphorylase kinase family signature PR01049: P794-R805	BLIMPS_PRINTS
					Protein_Kinase_Serine/Threonine: I654-I666	MOTIFS
					protein_kinase_tyrosine.profile: E634-G691	PROFILESCAN
15	7163416CD1	641	S107 S135 S165 S189 S248 S255 S276 S290 S332 S351 S429 S560 S624 T106 T107 T124 T212 T238 T24 T322 T46 T505 T580 T99	N288	Eukaryotic protein kinase domain pkinase: L407-Y601	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15					PROTEIN KINASE DOMAIN DM00004 | P35465 | 271-510: Y410-S628 DM00004 | I49376 | 270-509: K412-S628 DM00004 | Q03497 | 622-861: V411-S628 DM00004 | P50527 | 388-627: S409-S628	BLAST_DOMO
					KINASE SERINE/THREONINE PROTEIN TRANSFERASE ATP-BINDING PROTEIN PHOSPHORYLATION P21 ACTIVATED ACTIVATED HOMOLOG SYNDROME PD002852: I12-L44 (P=3.0e-06)	BLAST_PRODOM
					Tyrosine kinase catalytic domain signature PR00109: M481-S494, Y516-L534, G563-I573, V582-D604	BLIMPS_PRINTS
					Protein_Kinase_ATP I413-K436	MOTIFS
16	7472822CD1	576	S109 S136 S220 S255 S266 S31 S313 S318 S323 S327 S336 S451 S505 S506 S8 T152 T213 T333 T353 T364 T403 T447 T470 T497 T517 T557 Y440 Y482 Y59	N334	Guanylate kinase: T404-N500	HMMER_PFAM
					GUANYLATE KINASE DM00755 | A57653 | 370-570: P359-P570 DM00755 | I38757 | 709-898: R369-P570 DM00755 | S32545 | 1-196: R369-K556 DM00755 | P31007 | 765-954: R369-P570	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16					PROTEIN DOMAIN SH3 KINASE GUANYLATE TRANSFERASE ATP-BINDING REPEAT GMP MEMBRANE PD001338: T403-E496	BLAST_PRODOM
					PROTEIN SH3 DOMAIN PERIPHERAL PLASMA MEMBRANE CALMODULIN BINDING CASK CAMGUK CALCIUM PD008238: M1-I139	BLAST_PRODOM
					PROTEIN MAGUK P55 SUBFAMILY MEMBER DISCS LARGE HOMOLOG SH3 DOMAIN PD152180: K230-R297	BLAST_PRODOM
					PROTEIN MAGUK P55 SUBFAMILY MEMBER MPP3 DISCS LARGE HOMOLOG SH3 PD090357: S318-T403	BLAST_PRODOM
					Guanylate kinase protein BL00856: V400-I420, D428-R475	BLIMPS_BLOCKS
					SH3 domain signature PR00452: R284-R296, M231-P241, A252-Q267	BLIMPS_PRINTS
					PDZ domain (Also known as DHR or GLGF). PDZ: I139-G219	HMMER_PFAM
					SH3 domain SH3: M231-R296	HMMER_PFAM
					Guanylate_Kinase: T403-I420	MOTIFS
					signal_cleavage: M1-S31	SPSCAN

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	7477486CD1	794	S130 S158 S19 S201 S291 S327 S357 S379 S420 S443 S463 S512 S524 S571 S579 S602 S635 S659 S684 S692 S715 S731 S774 T145 T433 T488 T539 T591		PROTEIN KINASE DOMAIN DM00004|P34244|82-359: I71-S291 DM00004|JC1446|20-261: R51-L292 DM00004|P54645|17-258: L52-L292 DM00004|A53621|18-258: L52-L292	BLAST_DOMO
					KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: R42-L138, L144-A194 S209-F247, I270-L302	BLAST_PRODOM
					Tyrosine kinase catalytic domain signature PR00109: L126-V139, F162-L180, A228-D250, I270-L292	BLIMPS_PRINTS
					Eukaryotic protein kinase domain pkinase: Y50-Y301	HMMER_PFAM
					Protein_Kinase_ATP L56-K79	MOTIFS
					Protein_Kinase_Serine/Threonine: I168-L180	MOTIFS
					protein_kinase_tyrosine.profile: K120-S201	PROFILESCAN
18	3773709CD1	504	S117 S142 S152 S169 S232 S339 T274 T333 T375 T459 T6 T96 Y17	N131 N132 N178 N216	XYLULOKINASE DM02388|P18157|1-492: F20-M498	BLAST-DOMO
					GLYCEROL 3PHOSPHOTRANSFERASE GLYCEROKINASE GK PD001007: G239-A448	BLAST-PRODOM
					SIMILAR TO GLYCEROL KINASE PD130307: F20-K137	BLAST-PRODOM
					FGGY family of carbohydrate kinase proteins BL00933: F20-C43, Y54-P64, S159-N178, T212-V248, G414-L429	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18					FGGY family of carbohydrate kinases signatures prok_carb_kinases.prf: P350-K409	PROFILESCAN
					FGGY family of carbohydrate kinases FGGY: L92-R122, L172-D224, F238-A448	HMMER-PFAM
					Fggy_Kinases_2: A366-E386	MOTIFS
19	7477204CD1	553	S187 S23 S36 S380 S399 S544 S58 T138 T139 T213 T348 T407 T537 T79 T85	N418 N543	PROTEIN KINASE DOMAIN DM00004|P32298|157-401: F194-G440	BLAST-DOMO
					RECEPTOR KINASE PD001932: K455-N531	BLAST-PRODOM
					Regulator of G-protein PF00615: F163-K179 V267-I280	BLIMPS-PFAM
					Tyrosine kinase catalytic domain signature PR00109: F419-S441, M268-Y281, H306-L324, G352-L362, V372-Y394	BLIMPS-PRINTS
					GPCR kinase signature PR00717: Y172-Q184, K230-S248, P469-I486, V492-F505, N507-T524	BLIMPS-PRINTS
					Protein kinases signatures and profile protein_kinase_tyr.prf: R292-K345	PROFILESCAN
					Regulator of G protein signaling domain RGS: N55-P78, P162-L176	HMMER-PFAM
					Eukaryotic protein kinase domain pkinase: F191-F454	HMMER-PFAM
					Protein_Kinase_Atp: L197-K220	MOTIFS
					Protein_Kinase_St: I312-L324	MOTIFS

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>20</td><td>3016969CD1</td><td>871</td><td>S121 S123 S135 S153 S167 S203 S293 S33 S353 S409 S542 S557 S571 S597 S640 S652 S665 S667 S727 S81 T172 T417 T516 T526 T76 T844</td><td>N211</td><td>PROTEIN KINASE DOMAIN DM00004|S07571|5152-5396: Q580-P812</td><td>BLAST-DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine PR00109: Y684-I702, T751-E773, I581-A603</td><td>BLIMPS-PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Eukaryotic protein kinase domain pkinase: F575-L827</td><td>HMMER-PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein_Kinase_Tyr: I690-I702</td><td>MOTIFS</td></tr>
<tr><td>21</td><td>063497CD1</td><td>765</td><td>S162 S181 S259 S286 S291 S410 S431 S437 S472 S479 S495 S531 S539 S544 S550 S569 S576 S597 S639 S646 S661 S676 T172 T319 T365 T474 T478 T50 T543 T622 T623 T684 T714 T716</td><td>N219 N289 N588 N618</td><td>Eukaryotic protein kinase domain: Y16-L269</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature PR00109:L92-M105, Y129-F147, V238-L260</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>SNF1RELATED KINASE PD127501:Q346-D579 PD070820:T715-I765, E642-G693, I345-P370</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>ZK524.4 PROTEIN SNF1RELATED KINASE PD156028:I282-I345</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>KINASE TRANSFERASE ATP BINDING SERINE/ THREONINE PHOSPHORYLATION RECEPTOR TYROSINE TRANSMEMBRANE PD000001:L18-V145, V238-W268, G168-F215</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P27448|58-297:K20-L260
DM00004|I48609|55-294:K20-L260
DM00004|Q05512|55-294:K20-L260
DM00004|JC1446|20-261:L18-L260</td><td>BLAST_DOMO</td></tr>
</table>

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>21</td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
L22-K45</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
V135-F147</td><td>MOTIFS</td></tr>
<tr><td>22</td><td>1625436CD1</td><td>588</td><td>S109 S355 S356 S36 S427 S433 S51 S557 S79 T262 T383 T408 T409 T410 T47 T488 T94</td><td>N313 N394 N407 N424</td><td>Eukaryotic protein kinase domain:
Y14-V272</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases signatures and profile:
F85-E167</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:H126-L144</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>KINASE II CALCIUM/CALMODULIN DEPENDENT SUBUNIT TRANSFERASE SERINE/THREONINE
PD004250:E500-Q588
PD001779:R456-V499, V272-S329, T396-A417</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|P11798|15-261:L16-A263
DM00004|JU0270|16-262:E18-A263
DM00004|A44412|16-262:E18-A263
DM00004|S57347|21-266:L20-T262</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases ATP-binding region signature:
L20-K43</td><td>MOTIFS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
I132-L144</td><td>MOTIFS</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23	3330646CD1	1798	S74 S92 S1084 S108 S130 S1100 S166 S171 S1205 S200 S204 S1195 S230 S253 S1214 S281 S480 S1230 S503 S508 S1225 S533 S775 S1229 S806 S811 S1272 S817 S825 S1256 S846 S854 S1332 S860 S874 S1337 S909 S914 S1418 S931 S1425 S1429 S1447 S1459 S1491 S1503 S1504 S1541 S1650 S1657 S1660 S1671 S1698 S1717 S1771 T266 T506 T1014 T514 T565 T1036 T581 T729 T1040 T759 T786 T1117 T815 T82 T1189 T871 T916 T1236 T925 T949 T1244 T1424 T1480 T1675 T1765	N142 N1193 N1252 N1293	Eukaryotic protein kinase domain: F512-F785	HMMER_PFAM

Table 3 (cont.)

<table>
<tr><th>SEQ ID NO:</th><th>Incyte Polypeptide ID</th><th>Amino Acid Residues</th><th>Potential Phosphorylation Sites</th><th>Potential Glycosylation Sites</th><th>Signature Sequences, Domains and Motifs</th><th>Analytical Methods and Databases</th></tr>
<tr><td>23</td><td></td><td></td><td></td><td></td><td>PDZ domain:
P1104-L1191</td><td>HMMER_PFAM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Protein kinases signatures and profile:
F579-M659</td><td>PROFILESCAN</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Tyrosine kinase catalytic domain signature
PR00109:M589-K602, Y625-I643, V706-D728</td><td>BLIMPS_PRINTS</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>MICROTUBULE ASSOCIATED TESTIS SPECIFIC SERINE/THREONINE PROTEIN KINASE
PD142315:H1313-T1798
PD135564:V61-Y320, L1151-P1363
PD182663:E863-H1139</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE SERINE/THREONINE KIN4 MICROTUBULE ASSOCIATED TESTIS SPECIFIC
PD041650:K321-D511</td><td>BLAST_PRODOM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>PROTEIN KINASE DOMAIN
DM00004|A54602|455-712:T514-G772
DM08046|P05986|1-397: S508-K658, V685-E829, D268-P291
DM00004|S42867|75-498: I515-T666, H672-F813
DM00004|S42864|41-325: E513-K658, H672-T773</td><td>BLAST_DOMO</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td>Serine/Threonine protein kinases active-site signature:
I631-I643</td><td>MOTIFS</td></tr>
</table>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
24	3562763CD1	362	S123 S157 S25 S325 S81 T164 T197 T260 T280 T286 T324 T353	N110 N165	transmembrane domain: A263-D283	HMMER
					Eukaryotic protein kinase domain: Y30-L351	HMMER_PFAM
					Protein kinases signatures and profile: T164-G218	PROFILESCAN
					Tyrosine kinase catalytic domain signature PR00109: M143-L156, F178-I196, M326-A348	BLIMPS_PRINTS
					PROTEIN KINASE DOMAIN DM00004|Q02723|16-259: K111-V215, N232-V304 DM00004|A54602|455-712:N110-L316, I36-I61 DM00004|P23573|10-277: L139-K214, E35-L102, F248-A348 DM00004|A57459|417-662:Y138-S325, E35-L73	BLAST_DOMO
					Protein kinases ATP-binding region signature: I36-K59	MOTIFS
					Serine/Threonine protein kinases active-site signature: I184-I196	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	621293CD1	275			Adenylate kinase: L69-P205	HMMER_PFAM
					Adenylate kinase proteins. BL00113:L68-L84, N92-R135, C141-L155	BLIMPS_BLOCKS
					Adenylate kinase signature PR00094:L68-A81, G96-G110, W146-N162	BLIMPS_PRINTS
26	7480774CD1	660	S104 S106 S167 S199 S226 S325 S338 S339 S343 S355 S381 S458 S46 S629 S96 T117 T151 T160 T183 T210 T468 T500 T83 T90 T99	N177	INOSITOL 3 KINASE 1D MYOINOSITOL TRISPHOSPHATE 5 TRISPHOSPHATE IP3K IP3 TRANSFERASE KINASE CALMODULIN BINDING PD010031:Q446-Q659, P377-Q442	BLAST_PRODOM
					CALMODULIN-BINDING DOMAIN DM07435|P42335|210-672:E315-Q659 DM07435|P23677|1-461:G261-Q659	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
27	2011384CB1	822	282-377	6829315H1 (SINTNOR01)	44	743
				g2954208	1	282
28	2004888CB1	1376	1349-1376, 499-635	5545302T6 (TESTNOC01)	713	1376
				674588R6 (CRBLNOT01)	517	1256
				5562195F8 (BRSTDIT01)	1	644
29	2258952CB1	3468	1-983, 1461-1908, 3369-3468	3219989H1 (COLNNON03)	3223	3468
				2258952T6 (OVARTUT01)	2757	3353
				FL2258952_g7458755_000012_g3766209	33	2849
				7126256H1 (COLNDIY01)	2527	3076
				g1633937	2718	3385
				7677920H1 (NOSETUE01)	1	601
30	7473244CB1	2831	1-243, 834-1782	2660853T6 (LUNGTUT09)	2249	2831
				5216205F6 (BRSTNOT35)	1789	2681
				6854507F8 (BRAIFEN08)	763	1471
				55057226H1	354	1145
				5911008F6 (BRAIFEN05)	1299	1988
				2074751F6 (ISLTNOT01)	1626	2118
				6881535J1 (BRAHTDR03)	1	582
31	1242491CB1	2693	1-317, 2569-2693	70006068D1	1296	1838
				70006347D1	1162	1747
				7934296H1 (COLNDIS02)	2109	2693
				70003021D1	1740	2337
				7226035H1 (LUNGTMC01)	725	1187

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
31				5755513H1 (LUNGNOT35)	672	1102
				70004229D1	1874	2338
				55052947H1	1	694
32	2634875CB1	2973	1-1353, 2203-2560	4009430F6 (MUSCNOT10)	959	1432
				5168601H1 (MUSCDMT01)	1691	1965
				5672440H1 (MUSLTDT01)	2213	2414
				6903523H1 (MUSLTDR02)	1833	2344
				55052146J1	1475	1654
				6217472F6 (MUSCDIT06)	2263	2973
				3585116F6 (293TF4T01)	623	1126
				GBI.g7242443_000006 .edit	1059	1585
				55052619J1	1	807
				2634875H1 (BONTNOT01)	1521	1764
33	3951059CB1	2066	532-772, 1830-1886, 1966-2066	6882814J1 (BRAHTDR03)	1489	2066
				55058330J1	396	1316
				FL452484_00001	1	970
				71179403V1	1052	1745
34	7395890CB1	3975	1-326, 3951-3975, 2980-3355, 3666-3731, 1813-2074, 1066-1098	6771964H1 (BRAUNOR01)	715	1432
				6770122H1 (BRAUNOR01)	1471	2040
				6771964J1 (BRAUNOR01)	2028	2713
				7393659H1 (BRABDIE02)	186	799
				55052405H1	1	218
				2570554R6 (HIPOAZT01)	2495	3012
				7660364H1 (OVARNOE02)	1861	2459
				FL034583_00001	2778	3584

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
34				7395271H1 (BRABDIE02)	256	896
				6200064H1 (PITUNON01)	2715	3162
				7395911H1 (BRABDIE02)	896	1481
				GNN.g8439948_000007.edit2.comp	3181	3975
				6873077H1 (BRAGNON02)	1327	1999
35	7475546CB1	1918	1-46, 658-1061	6623984J1 (UTRMTMR02)	655	1287
				7192851H2 (BRATDIC01)	497	1107
				6810083J1 (SKIRNOR01)	1254	1918
				7013748H1 (KIDNNOC01)	1	580
36	7477076CB1	1689	1-66	7190770H1 (BRATDIC01)	216	771
				55051332H1	1	282
				6819441H1 (OVARDIR01)	1077	1689
				7758313J1 (SPLNTUE01)	558	922
				GNN:g807680_edit	820	1476
37	1874092CB1	1054	1-30	1874092F6 (LEUKNOT02)	604	1054
				7315561H1 (SYNODIN02)	1	633
38	4841542CB1	3360	1-172, 2484-2523, 650-1457, 2247-2417	71224917V1	2797	3360
				70858292V1	2345	3032
				8045106H1 (OVARTUE01)	1719	2379
				7617315J1 (KIDNTUE01)	1036	1632
				7609838J1 (KIDCTME01)	783	1346
				70856122V1	2494	3142
				71225608V1	1597	2126
				55053856H1	1	826

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
39	7472695CB1	2240	1-20, 101-131, 704-1001	7191541F6 (BRATDIC01)	1	906
				71872279V1	911	1501
				4211726T8 (BRONDIT01)	1466	2181
				71870527V1	1717	2240
				71870095V1	669	1374
				2013786T6 (TESTNOT03)	1551	2217
40	7477966CB1	3340	1-980, 1504-1710, 3315-3340	1513994T6 (PANCTUT01)	2768	3340
				6802962H1 (COLENOR03)	2241	2824
				55052773H1	1376	2254
				1513994F6 (PANCTUT01)	2155	2776
				55052765H1	894	1745
				7607337J1 (COLRTUE01)	594	1258
				6802518H1 (COLENOR03)	551	858
				7677920H1 (NOSETUE01)	1	598
41	7163416CB1	2539	1-228, 913-1225, 1994-2539	7715351J1 (SINTFEE02)	1	649
				1625532H1 (COLNPOT01)	1779	1993
				7163416F8 (PLACNOR01)	1888	2539
				7701682J1 (PENHTUE02)	815	1434
				7715351H1 (SINTFEE02)	399	1037
				7077243H1 (BRAUTDR04)	1306	1979
42	7472822CB1	2377	2341-2377, 1093-1463, 1625-2081	71982976V1	913	1546
				71983661V1	793	1520
				71986606V1	1494	2168
				55052941J1	1	886
				71983943V1	1551	2193
				71983660V1	1642	2377

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
43	7477486CB1	2897	2698-2763, 1-365, 2314-2623, 1516-1614, 2804-2897	4029722F8 (BRAINOT23)	2042	2584
				6910737R6 (PITUDIR01)	462	1370
				7237528H1 (BRAINOY02)	2348	2897
				7674962H2 (NOSETUE01)	125	589
				71982594V1	1386	1991
				6629715R6 (HEALDIR01)	637	1476
				GNN.g6165121_004.edit	1	506
				6950253H1 (BRAITDR02)	1480	2176
44	3773709CB1	3361	1-168, 1479-1982, 3336-3361	6938382F6 (FTUBTUR01)	116	850
				4383108H1 (BRAVUTT02)	1	257
				7365206H1 (OVARDIC01)	2019	2580
				55024481H1 (PKINDNV08)	791	1462
				4119492H1 (BRSTTUT25)	3104	3361
				70783206V1	1969	2579
				3432983T6 (SKINNOT04)	2555	3217
				70782455V1	1361	2005
				70143324V1	2631	3219
				70784860V1	1463	2006
45	7477204CB1	1662	854-1662, 1-807	GNN.g8139716_edit	1	1662
46	3016969CB1	3225	1-916, 1154-1362, 3144-3225	71873834V1	1555	2122
				5751549F8 (LUNGNOT35)	2153	2740
				7718401J1 (SINTFEE02)	1341	2100
				7354408H1 (HEARNON03)	2779	3225
				71872969V1	1969	2707
				71875134V1	885	1440

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
46				3016969T6 (MUSCNOT07)	2532	3211
				6200811F6 (PITUNON01)	808	1403
				55052669H1	1	852
47	063497CB1	4772	1-431, 4420-4540, 2098-2130, 3522-3599, 2875-3036	6581829H1 (HEACDIC01)	2823	3464
				7199634H1 (LUNGFER04)	602	1153
				6936880H1 (FTUBTUR01)	3000	3714
				1449223H1 (PLACNOT02)	4029	4248
				4787168H1 (BRATNOT03)	3705	3964
				7714789H1 (SINTFEE02)	1198	1849
				7714789J1 (SINTFEE02)	4189	4772
				063497H1 (PLACNOB01)	1661	1880
				8025257J1 (ENDMUNE01)	1	702
				7381417H1 (ENDMUNE01)	1790	2359
				4351289H1 (CONFTMT01)	3884	4222
				5068175H1 (PANCNOT23)	3675	3946
				7380657H1 (ENDMUNE01)	772	1305
				4051307H1 (SINTNOT18)	2689	2972
				7627517J1 (GBLADIE01)	2393	2919
				7629590H1 (GBLADIE01)	1953	2559
48	1625436CB1	1880	948-1167	5772228H1 (BRAINOT20)	844	1420
				72285173V1	673	1148

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
48				7353062H1 (HEARNON03)	1	610
				7154515H1 (BRAMNOA01)	1164	1839
				6764194H1 (BRAUNOR01)	1370	1880
				72284772V1	491	1135
49	3330646CB1	5747	1-1738, 2291-2733, 3677-4763	8178538H2 (EYERNON01)	5053	5722
				7218734H1 (COLNTMC01)	4882	5570
				8013776H1 (HEARNOC04)	4245	4904
				8006864H1 (PENIFEC01)	442	1064
				7711762H2 (TESTTUE02)	688	1292
				55124907H1	1301	2151
				8009629H1 (NOSEDIC02)	3681	4314
				7054991H1 (BRALNON02)	5099	5747
				55124907J1	1250	2101
				8267426H1 (MIXDUNF03)	2739	3511
				8054655J1 (ESOGTUE01)	2905	3529
				7930953H1 (COLNDIS02)	4339	4966
				7978939H1 (LSUBDMC01)	1	504
				7719236J1 (SINTFEE02)	2085	2746
				60215898V1	2234	2776
				6779321J1 (OVARDIR01)	3439	4230
50	3562763CB1	3418	1564-1627, 1-376, 975-1073, 3066-3418	55053205H1	523	1210
				7321924H1 (NOSETUE01)	1843	2392
				7278180H1 (BMARTXE01)	2873	3418

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
50				400518R6 (PITUNOT02)	873	1430
				6816641J1 (ADRETUR01)	1297	1981
				g2963935	1	383
				55143790J1	2257	3143
				55067380J2	314	579
				55143774J1	2577	3148
51	621293CB1	995	1-372, 410-468	72335268V1	1	508
				71870548V1	477	994
52	7480774CB1	2459	1664-2459, 1-110	71440281V1	685	1345
				71438714V1	652	1226
				7082565H1 (STOMTMR02)	1	688
				71432228V1	1798	2459
				71431941V1	1257	1972
				6472388H1 (PLACFEB01)	1352	1985

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
27	2011384CB1	SINTNOR01
28	2004888CB1	TESTNOT03
29	2258952CB1	COLENOR03
30	7473244CB1	ISLTNOT01
31	1242491CB1	LUNGNOT02
32	2634875CB1	MUSCNOT07
33	3951059CB1	DRGCNOT01
34	7395890CB1	BRABDIE02
35	7475546CB1	CORPNOT02
36	7477076CB1	BRATDIC01
37	1874092CB1	LEUKNOT02
38	4841542CB1	KIDNNOT05
39	7472695CB1	TESTNOT03
40	7477966CB1	COLENOR03
41	7163416CB1	ESOGTME01
42	7472822CB1	BRABDIR03
43	7477486CB1	BRAITDR03
44	3773709CB1	SINTNOR01
46	3016969CB1	COLNNOT41
47	063497CB1	ENDMUNE01
48	1625436CB1	BRACNOK02
49	3330646CB1	HNT2AGT01
50	3562763CB1	BRAHNOE01
51	621293CB1	KIDNNOT09
52	7480774CB1	BLADTUT02

Table 6

Library	Vector	Library Description
BLADTUT02	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology indicated grade 3 invasive transitional cell carcinoma. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
BRABDIE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day, for 40 years).
BRABDIR03	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day for 40 years).
BRACNOK02	PSPORT1	This amplified and normalized library was constructed using RNA isolated from posterior cingulate tissue removed from an 85-year-old Caucasian female who died from myocardial infarction and retroperitoneal hemorrhage. Pathology indicated atherosclerosis, moderate to severe, involving the circle of Willis, middle cerebral, basilar and vertebral arteries; infarction, remote, left dentate nucleus; and amyloid plaque deposition consistent with age. There was mild to moderate leptomeningeal fibrosis, especially over the convexity of the frontal lobe. There was mild generalized atrophy involving all lobes. The white matter was mildly thinned. Cortical thickness in the temporal lobes, both maximal and minimal, was slightly reduced. The substantia nigra pars compacta appeared mildly depigmented. Patient history included COPD, hypertension, and recurrent deep venous thrombosis. 6.4 million independent clones from this amplified library were normalized in one round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791.
BRAHNOE01	pINCY	Library was constructed RNA isolated from posterior hippocampus tissue removed from a 45-year-old Caucasian female who died from a dissecting aortic aneurysm and ischemic bowel disease. Pathology indicated mild arteriosclerosis involving the cerebral cortical white matter and basal ganglia. Grossly, there was mild meningeal fibrosis and mild focal atherosclerotic plaque in the middle cerebral artery, as well as vertebral arteries bilaterally. Microscopically, the cerebral hemispheres, brain stem and cerebellum reveal focal areas in the white matter that contain blood vessels that were barrel-shaped, hyalinized, with hemosiderin-laden macrophages in the Virchow-Robin space. In addition, there were scattered neurofibrillary tangles within the basolateral nuclei of the amygdala. Patient

Table 6 (cont.)

Library	Vector	Library Description
		history included mild atheromatosis of aorta and coronary arteries, bowel and liver infarct due to aneurysm, physiologic fatty liver associated with obesity, mild diffuse emphysema, thrombosis of mesenteric and portal veins, cardiomegaly due to hypertrophy of left ventricle, arterial hypertension, acute pulmonary edema, splenomegaly, obesity (300 lb.), leiomyoma of uterus, sleep apnea, and iron deficiency anemia.
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydorthorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRATDIC01	pINCY	This large size-fractionated library was constructed using RNA isolated from diseased brain tissue removed from the left temporal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology for the left temporal lobe, including the mesial temporal structures, indicated focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. The left frontal lobe showed a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. The frontal lobe tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, depressive disorder, and tobacco abuse in remission. Previous surgeries included tendon transfer. Patient medications included minocycline hydrochloride, Tegretol, phenobarbital, vitamin C, Pepcid, and Pevaryl. Family history included brain cancer in the father.
COLENOR03	PCDNA2.1	Library was constructed using RNA isolated from colon epithelium tissue removed from a 13-year-old Caucasian female who died from a motor vehicle accident.
COLNNOT41	pINCY	Library was constructed using RNA isolated from colon tissue removed from a 37-year-old female during a partial gastrojejunectomy. Pathology indicated a portion

Table 6 (cont.)

Library	Vector	Library Description
		of stomach and jejunum with an intact anastomotic site. The stomach showed a mild chronic gastritis without helicobacter pylori organisms. Normal appearing submucosal and myenteric plexus ganglion cells were noted. The jejunum had no significant abnormality.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
DRGCNOT01	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
ENDMUNE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from untreated umbilical artery endothelial cell tissue removed from a Caucasian male (Clonetics) newborn.
ESOGTME01	PSPORT1	This 5' biased random primed library was constructed using RNA isolated from esophageal tissue removed from a 53-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, and regional lymph node biopsy. Pathology indicated no significant abnormality in the non-neoplastic esophagus. Pathology for the matched tumor tissue indicated invasive grade 4 (of 4) adenocarcinoma, forming a sessile mass situated in the lower esophagus, 2 cm from the gastroesophageal junction and 7 cm from the proximal margin. The tumor invaded through the muscularis propria into the adventitial soft tissue. Metastatic carcinoma was identified in 2 of 5 paragastric lymph nodes with perinodal extension. The patient presented with dysphagia. Patient history included membranous nephritis, hyperlipidemia, benign hypertension, and anxiety state. Previous surgeries included an adenotonsillectomy, appendectomy, and inguinal hernia repair. The patient was not taking any medications. Family history included atherosclerotic coronary artery disease, alcoholic cirrhosis, alcohol abuse, and an abdominal aortic aneurysm rupture in the father; breast cancer in the mother; a myocardial infarction and atherosclerotic coronary artery disease in the sibling(s); and myocardial infarction and atherosclerotic coronary artery disease in the grandparent(s).
HNT2AGT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937233), using RNA isolated from the

Table 6 (cont.)

Library	Vector	Library Description
		hNT2 cell line derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor. Cells were treated with retinoic acid for 5 weeks and with mitotic inhibitors for two weeks and allowed to mature for an additional 4 weeks in conditioned medium.
ISLTNOT01	pINCY	Library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.
KIDNNOT05	PSPORT1	Library was constructed using RNA isolated from the kidney tissue of a 2-day-old Hispanic female, who died from cerebral anoxia. Family history included congenital heart disease.
KIDNNOT09	pINCY	Library was constructed using RNA isolated from the kidney tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
LEUKNOT02	pINCY	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
LUNGNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from the lung tissue of a 47-year-old Caucasian male, who died of a subarachnoid hemorrhage.
MUSCNOT07	pINCY	Library was constructed using RNA isolated from muscle tissue removed from the forearm of a 38-year-old Caucasian female during a soft tissue excision. Pathology for the associated tumor tissue indicated intramuscular hemangioma. Family history included breast cancer, benign hypertension, cerebrovascular disease, colon cancer, and type II diabetes.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
TESTNOT03	PBLUESCRIPT	Library was constructed using RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-26,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-26.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:27-52.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-52,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

18. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of PKIN in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of PKIN in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-26.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.

74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.

75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.

76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.

77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.

78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.

79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.

80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.

81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.

82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.

83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.

84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.

85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.

86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.

87. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:43.

88. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:44.

89. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:45.

90. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:46.

91. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:47.

92. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:48.

93. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:49.

94. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:50.

95. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:51.

96. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:52.

<110> INCYTE GENOMICS, INC.
YUE, Henry
LAL, Preeti
BANDMAN, Olga
BOROWSKY, Mark L.
AU-YOUNG, Janice
LU, Yan
GANDHI, Ameena R.
TRIBOULEY, Catherine M.
WALIA, Narinder K.
YAO, Monique G.
LU, Dyung Aina M.
GREENWALD, Sara R.
RAMKUMAR, Jayalaxmi
GRIFFIN, Jennifer A.
KEARNEY, Liam
BURFORD, Neil
NGUYEN, Danniel B.
TANG, Y. Tom
BAUGHN, Mariah R.
HE, Ann
THORNTON, Michael
HAFALIA, April
PATTERSON, Chandra
GURURAJAN, Rajagopal
LO, Terence P.
KHAH, Farrah A.
RECIPON, Shirley A.
AZIMZAI, Yalda
POLICKY, Jennifer L.
DING, Li
GRETHER, Megan
ELLIOTT, Vicki S.
THANGAVELU, Kavitha
BATRA, Sajeev
ISON, Craig H.

<120> HUMAN KINASES

<130> PI-0125 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/212,073; 60/213,467; 60/215,651; 60/216,605; 60/218,372; 60/228,056
<151> 2000-06-15; 2000-06-23; 2000-06-30; 2000-07-07; 2000-07-13; 2000-08-25

<160> 52

<170> PERL Program

<210> 1
<211> 273
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2011384CD1

<400> 1
Met Ser Gly Asp Lys Leu Leu Ser Glu Leu Gly Tyr Lys Leu Gly
1 5 10 15
Arg Thr Ile Gly Glu Gly Ser Tyr Ser Lys Val Lys Val Ala Thr

```
                    20                  25                  30
Ser Lys Lys Tyr Lys Gly Thr Val Ala Ile Lys Val Val Asp Arg
                    35                  40                  45
Arg Arg Ala Pro Pro Asp Phe Val Asn Lys Phe Leu Pro Arg Glu
                    50                  55                  60
Leu Ser Ile Leu Arg Gly Val Arg His Pro His Ile Val His Val
                    65                  70                  75
Phe Glu Phe Ile Glu Val Cys Asn Gly Lys Leu Tyr Ile Val Met
                    80                  85                  90
Glu Ala Ala Ala Thr Asp Leu Leu Gln Ala Val Gln Arg Asn Gly
                    95                  100                 105
Arg Ile Pro Gly Val Gln Ala Arg Asp Leu Phe Ala Gln Ile Ala
                    110                 115                 120
Gly Ala Val Arg Tyr Leu His Asp His His Leu Val His Arg Asp
                    125                 130                 135
Leu Lys Cys Glu Asn Val Leu Leu Ser Pro Asp Glu Arg Arg Val
                    140                 145                 150
Lys Leu Thr Asp Phe Gly Phe Gly Arg Gln Ala His Gly Tyr Pro
                    155                 160                 165
Asp Leu Ser Thr Thr Tyr Cys Gly Ser Ala Ala Tyr Ala Ser Pro
                    170                 175                 180
Glu Val Leu Leu Gly Ile Pro Tyr Asp Pro Lys Lys Tyr Asp Val
                    185                 190                 195
Trp Ser Met Gly Val Val Leu Tyr Val Met Val Thr Gly Cys Met
                    200                 205                 210
Pro Phe Asp Asp Ser Asp Ile Ala Gly Leu Pro Arg Arg Gln Lys
                    215                 220                 225
Arg Gly Val Leu Tyr Pro Glu Gly Leu Glu Leu Ser Glu Arg Cys
                    230                 235                 240
Lys Ala Leu Ile Ala Glu Leu Leu Gln Phe Ser Pro Ser Ala Arg
                    245                 250                 255
Pro Ser Ala Gly Gln Val Ala Arg Asn Cys Trp Leu Arg Ala Gly
                    260                 265                 270
Asp Ser Gly


<210> 2
<211> 329
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2004888CD1

<400> 2
Met Leu Thr Ser Leu Ala Gln Lys Trp Phe Pro Glu Leu Pro Leu
  1               5                  10                  15
Leu His Pro Glu Ile Gly Leu Leu Lys Tyr Met Asn Ser Gly Gly
                    20                  25                  30
Leu Leu Thr Met Ser Leu Glu Arg Asp Leu Leu Asp Ala Glu Pro
                    35                  40                  45
Met Lys Glu Leu Ser Ser Lys Arg Pro Leu Val Arg Ser Glu Val
                    50                  55                  60
Asn Gly Gln Ile Ile Leu Leu Lys Gly Tyr Ser Val Asp Val Asp
                    65                  70                  75
Thr Glu Ala Lys Val Ile Glu Arg Ala Ala Thr Tyr His Arg Ala
                    80                  85                  90
Trp Arg Glu Ala Glu Gly Asp Ser Gly Leu Leu Pro Leu Ile Phe
                    95                  100                 105
Leu Phe Leu Cys Lys Ser Asp Pro Met Ala Tyr Leu Met Val Pro
                    110                 115                 120
Tyr Tyr Pro Arg Ala Asn Leu Asn Ala Val Gln Ala Asn Met Pro
                    125                 130                 135
Leu Asn Ser Glu Glu Thr Leu Lys Val Met Lys Gly Val Ala Gln
                    140                 145                 150
Gly Leu His Thr Leu His Lys Ala Asp Ile Ile His Gly Ser Leu
```

```
                    155                                 160                                 165
His Gln Asn Asn Val Phe Ala Leu Asn Arg Glu Gln Gly Ile Val
                    170                                 175                                 180
Gly Asp Phe Asp Phe Thr Lys Ser Val Ser Gln Arg Ala Ser Val
                    185                                 190                                 195
Asn Met Met Val Gly Asp Leu Ser Leu Met Ser Pro Glu Leu Lys
                    200                                 205                                 210
Met Gly Lys Pro Ala Ser Pro Gly Ser Asp Leu Tyr Ala Tyr Gly
                    215                                 220                                 225
Cys Leu Leu Leu Trp Leu Ser Val Gln Asn Gln Glu Phe Glu Ile
                    230                                 235                                 240
Asn Lys Asp Gly Ile Pro Lys Val Asp Gln Phe His Leu Asp Asp
                    245                                 250                                 255
Lys Val Lys Ser Leu Leu Cys Ser Leu Ile Cys Tyr Arg Ser Ser
                    260                                 265                                 270
Met Thr Ala Glu Gln Val Leu Asn Ala Glu Cys Phe Leu Met Pro
                    275                                 280                                 285
Lys Glu Gln Ser Val Pro Asn Pro Glu Lys Asp Thr Glu Tyr Thr
                    290                                 295                                 300
Leu Tyr Lys Lys Glu Glu Glu Ile Lys Thr Glu Asn Leu Asp Lys
                    305                                 310                                 315
Cys Met Glu Lys Thr Arg Asn Gly Glu Ala Asn Phe Asp Cys
                    320                                 325
```

<210> 3
<211> 938
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2258952CD1

<400> 3

```
Met Met Ser Asp Thr Ser Thr Phe Pro Asn His Pro Ser Ser Pro
1                   5                   10                  15
Ala Ala Ser Pro Ser Gly Gly Arg Gly Val Met Ala Ser Pro Ala
                    20                  25                  30
Trp Asp Arg Ser Lys Gly Trp Ser Gln Thr Pro Gln Arg Ala Asp
                    35                  40                  45
Phe Val Ser Thr Pro Leu Gln Val His Thr Leu Arg Pro Glu Asn
                    50                  55                  60
Leu Leu Leu Val Ser Thr Leu Asp Gly Ser Leu His Ala Leu Ser
                    65                  70                  75
Lys Gln Thr Gly Asp Leu Lys Trp Thr Leu Arg Asp Asp Pro Val
                    80                  85                  90
Ile Glu Gly Pro Met Tyr Val Thr Glu Met Ala Phe Leu Ser Asp
                    95                  100                 105
Pro Ala Asp Gly Ser Leu Tyr Ile Leu Gly Thr Gln Lys Gln Gln
                    110                 115                 120
Gly Leu Met Lys Leu Pro Phe Thr Ile Pro Glu Leu Val His Ala
                    125                 130                 135
Ser Pro Cys Arg Ser Ser Asp Gly Val Phe Tyr Thr Gly Arg Lys
                    140                 145                 150
Gln Asp Ala Trp Phe Val Val Asp Pro Glu Ser Gly Glu Thr Gln
                    155                 160                 165
Met Thr Leu Thr Thr Glu Gly Pro Ser Thr Pro Arg Leu Tyr Ile
                    170                 175                 180
Gly Arg Thr Gln Tyr Thr Val Thr Met His Asp Pro Arg Ala Pro
                    185                 190                 195
Ala Leu Arg Trp Asn Thr Thr Tyr Arg Arg Tyr Ser Ala Pro Pro
                    200                 205                 210
Met Asp Gly Ser Pro Gly Lys Tyr Met Ser His Leu Ala Ser Cys
                    215                 220                 225
Gly Met Gly Leu Leu Leu Thr Val Asp Pro Gly Ser Gly Thr Val
                    230                 235                 240
Leu Trp Thr Gln Asp Leu Gly Val Pro Val Met Gly Val Tyr Thr
```

245 250 255
Trp His Gln Asp Gly Leu Arg Gln Leu Pro His Leu Thr Leu Ala
260 265 270
Arg Asp Thr Leu His Phe Leu Ala Leu Arg Trp Gly His Ile Arg
275 280 285
Leu Pro Ala Ser Gly Pro Arg Asp Thr Ala Thr Leu Phe Ser Thr
290 295 300
Leu Asp Thr Gln Leu Leu Met Thr Leu Tyr Val Gly Lys Asp Glu
305 310 315
Thr Gly Phe Tyr Val Ser Lys Ala Leu Val His Thr Gly Val Ala
320 325 330
Leu Val Pro Arg Gly Leu Thr Leu Ala Pro Ala Asp Gly Pro Thr
335 340 345
Thr Asp Glu Val Thr Leu Gln Val Ser Gly Glu Arg Glu Gly Ser
350 355 360
Pro Ser Thr Ala Val Arg Tyr Pro Ser Gly Ser Val Ala Leu Pro
365 370 375
Ser Gln Trp Leu Leu Ile Gly His His Glu Leu Pro Pro Val Leu
380 385 390
His Thr Thr Met Leu Arg Val His Pro Thr Leu Gly Ser Gly Thr
395 400 405
Ala Glu Thr Arg Pro Pro Glu Asn Thr Gln Ala Pro Ala Phe Phe
410 415 420
Leu Glu Leu Leu Ser Leu Ser Arg Glu Lys Leu Trp Asp Ser Glu
425 430 435
Leu His Pro Glu Glu Lys Thr Pro Asp Ser Tyr Leu Gly Leu Gly
440 445 450
Pro Gln Asp Leu Leu Ala Ala Ser Leu Thr Ala Val Leu Leu Gly
455 460 465
Gly Trp Ile Leu Phe Val Met Arg Gln Gln Gln Glu Thr Pro Leu
470 475 480
Ala Pro Ala Asp Phe Ala His Ile Ser Gln Asp Ala Gln Ser Leu
485 490 495
His Ser Gly Ala Ser Arg Arg Ser Gln Lys Arg Leu Gln Ser Pro
500 505 510
Ser Pro Glu Ser Pro Pro Ser Ser Pro Pro Ala Glu Gln Leu Thr
515 520 525
Val Val Gly Lys Ile Ser Phe Asn Pro Lys Asp Val Leu Gly Arg
530 535 540
Gly Ala Gly Gly Thr Phe Val Phe Arg Gly Gln Phe Glu Gly Arg
545 550 555
Ala Val Ala Val Lys Arg Leu Leu Arg Glu Cys Phe Gly Leu Val
560 565 570
Arg Arg Glu Val Gln Leu Leu Gln Glu Ser Asp Arg His Pro Asn
575 580 585
Val Leu Arg Tyr Phe Cys Thr Glu Arg Gly Pro Gln Phe His Tyr
590 595 600
Ile Ala Leu Glu Leu Cys Arg Ala Ser Leu Gln Glu Tyr Val Glu
605 610 615
Asn Pro Asp Leu Asp Arg Gly Gly Leu Glu Pro Glu Val Val Leu
620 625 630
Gln Gln Leu Met Ser Gly Leu Ala His Leu His Ser Leu His Ile
635 640 645
Val His Arg Asp Leu Lys Pro Gly Asn Ile Leu Ile Thr Gly Pro
650 655 660
Asp Ser Gln Gly Leu Gly Arg Val Val Leu Ser Asp Phe Gly Leu
665 670 675
Cys Lys Lys Leu Pro Ala Gly Arg Cys Ser Phe Ser Leu His Ser
680 685 690
Gly Ile Pro Gly Thr Glu Gly Trp Met Ala Pro Glu Leu Leu Gln
695 700 705
Leu Leu Pro Pro Asp Ser Pro Thr Ser Ala Val Asp Ile Phe Ser
710 715 720
Ala Gly Cys Val Phe Tyr Tyr Val Leu Ser Gly Gly Ser His Pro
725 730 735
Phe Gly Asp Ser Leu Tyr Arg Gln Ala Asn Ile Leu Thr Gly Ala
740 745 750

```
Pro Cys Leu Ala His Leu Glu Glu Glu Val His Asp Lys Val Val
                755                 760                 765
Ala Arg Asp Leu Val Gly Ala Met Leu Ser Pro Leu Pro Gln Pro
                770                 775                 780
Arg Pro Ser Ala Pro Gln Val Leu Ala His Pro Phe Phe Trp Ser
                785                 790                 795
Arg Ala Lys Gln Leu Gln Phe Phe Gln Asp Val Ser Asp Trp Leu
                800                 805                 810
Glu Lys Glu Ser Glu Gln Glu Pro Leu Val Arg Ala Leu Glu Ala
                815                 820                 825
Gly Gly Cys Ala Val Val Arg Asp Asn Trp His Glu His Ile Ser
                830                 835                 840
Met Pro Leu Gln Thr Asp Leu Arg Lys Phe Arg Ser Tyr Lys Gly
                845                 850                 855
Thr Ser Val Arg Asp Leu Leu Arg Ala Val Arg Asn Lys Lys His
                860                 865                 870
His Tyr Arg Glu Leu Pro Val Glu Val Arg Gln Ala Leu Gly Gln
                875                 880                 885
Val Pro Asp Gly Phe Val Gln Tyr Phe Thr Asn Arg Phe Pro Arg
                890                 895                 900
Leu Leu Leu His Thr His Arg Ala Met Arg Ser Cys Ala Ser Glu
                905                 910                 915
Ser Leu Phe Leu Pro Tyr Tyr Pro Pro Asp Ser Glu Ala Arg Arg
                920                 925                 930
Pro Cys Pro Gly Ala Thr Gly Arg
                935

<210> 4
<211> 795
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473244CD1

<400> 4
Met Ser Ala Arg Thr Pro Leu Pro Thr Val Asn Glu Arg Asp Thr
  1               5                  10                  15
Glu Asn His Thr Ser Val Asp Gly Tyr Thr Glu Pro His Ile Gln
                 20                  25                  30
Pro Thr Lys Ser Ser Ser Arg Gln Asn Ile Pro Arg Cys Arg Asn
                 35                  40                  45
Ser Ile Thr Ser Ala Thr Asp Glu Gln Pro His Ile Gly Asn Tyr
                 50                  55                  60
Arg Leu Gln Lys Thr Ile Gly Lys Gly Asn Phe Ala Lys Val Lys
                 65                  70                  75
Leu Ala Arg His Val Leu Thr Gly Arg Glu Val Ala Val Lys Ile
                 80                  85                  90
Ile Asp Lys Thr Gln Leu Asn Pro Thr Ser Leu Gln Lys Leu Phe
                 95                 100                 105
Arg Glu Val Arg Ile Met Lys Ile Leu Asn His Pro Asn Ile Val
                110                 115                 120
Lys Leu Phe Glu Val Ile Glu Thr Glu Lys Thr Leu Tyr Leu Val
                125                 130                 135
Met Glu Tyr Ala Ser Gly Gly Glu Val Phe Asp Tyr Leu Val Ala
                140                 145                 150
His Gly Arg Met Lys Glu Lys Glu Ala Arg Ala Lys Phe Arg Gln
                155                 160                 165
Ile Val Ser Ala Val Gln Tyr Cys His Gln Lys Tyr Ile Val His
                170                 175                 180
Arg Asp Leu Lys Ala Glu Asn Leu Leu Leu Asp Gly Asp Met Asn
                185                 190                 195
Ile Lys Ile Ala Asp Phe Gly Phe Ser Asn Glu Phe Thr Val Gly
                200                 205                 210
Asn Lys Leu Asp Thr Phe Cys Gly Ser Pro Pro Tyr Ala Ala Pro
                215                 220                 225
```

Glu Leu Phe Gln Gly Lys Lys Tyr Asp Gly Pro Glu Val Asp Val
230 235 240
Trp Ser Leu Gly Val Ile Leu Tyr Thr Leu Val Ser Gly Ser Leu
245 250 255
Pro Phe Asp Gly Gln Asn Leu Lys Glu Leu Arg Glu Arg Val Leu
260 265 270
Arg Gly Lys Tyr Arg Ile Pro Phe Tyr Met Ser Thr Asp Cys Glu
275 280 285
Asn Leu Leu Lys Lys Leu Leu Val Leu Asn Pro Ile Lys Arg Gly
290 295 300
Ser Leu Glu Gln Ile Met Lys Asp Arg Trp Met Asn Val Gly His
305 310 315
Glu Glu Glu Glu Leu Lys Pro Tyr Thr Glu Pro Asp Pro Asp Phe
320 325 330
Asn Asp Thr Lys Arg Ile Asp Ile Met Val Thr Met Gly Phe Ala
335 340 345
Arg Asp Glu Ile Asn Asp Ala Leu Ile Asn Gln Lys Tyr Asp Glu
350 355 360
Val Met Ala Thr Tyr Ile Leu Leu Gly Arg Lys Pro Pro Glu Phe
365 370 375
Glu Gly Gly Glu Ser Leu Ser Ser Gly Asn Leu Cys Gln Arg Ser
380 385 390
Arg Pro Ser Ser Asp Leu Asn Asn Ser Thr Leu Gln Ser Pro Ala
395 400 405
His Leu Lys Val Gln Arg Ser Ile Ser Ala Asn Gln Lys Gln Arg
410 415 420
Arg Phe Ser Asp His Ala Gly Pro Ser Ile Pro Pro Ala Val Ser
425 430 435
Tyr Thr Lys Arg Pro Gln Ala Asn Ser Val Glu Ser Glu Gln Lys
440 445 450
Glu Glu Trp Asp Lys Asp Val Ala Arg Lys Leu Gly Ser Thr Thr
455 460 465
Val Gly Ser Lys Ser Glu Met Thr Ala Ser Pro Leu Val Gly Pro
470 475 480
Glu Arg Lys Lys Ser Ser Thr Ile Pro Ser Asn Asn Val Tyr Ser
485 490 495
Gly Gly Ser Met Ala Arg Arg Asn Thr Tyr Val Cys Glu Arg Thr
500 505 510
Thr Asp Arg Tyr Val Ala Leu Gln Asn Gly Lys Asp Ser Ser Leu
515 520 525
Thr Glu Met Ser Val Ser Ser Ile Ser Ser Ala Gly Ser Ser Val
530 535 540
Ala Ser Ala Val Pro Ser Ala Arg Pro Arg His Gln Lys Ser Met
545 550 555
Ser Thr Ser Gly His Pro Ile Lys Val Thr Leu Pro Thr Ile Lys
560 565 570
Asp Gly Ser Glu Ala Tyr Arg Pro Gly Thr Thr Gln Arg Val Pro
575 580 585
Ala Ala Ser Pro Ser Ala His Ser Ile Ser Thr Ala Thr Pro Asp
590 595 600
Arg Thr Arg Phe Pro Arg Gly Ser Ser Ser Arg Ser Thr Phe His
605 610 615
Gly Glu Gln Leu Arg Glu Arg Arg Ser Val Ala Tyr Asn Gly Pro
620 625 630
Pro Ala Ser Pro Ser His Glu Thr Gly Ala Phe Ala His Ala Arg
635 640 645
Arg Gly Thr Ser Thr Gly Ile Ile Ser Lys Ile Thr Ser Lys Phe
650 655 660
Val Arg Arg Asp Pro Ser Glu Gly Glu Ala Ser Gly Arg Thr Asp
665 670 675
Thr Ser Arg Ser Thr Ser Gly Glu Pro Lys Glu Arg Asp Lys Glu
680 685 690
Glu Gly Lys Asp Ser Lys Pro Arg Ser Leu Arg Phe Thr Trp Ser
695 700 705
Met Lys Thr Thr Ser Ser Met Asp Pro Asn Asp Met Met Arg Glu
710 715 720
Ile Arg Lys Val Leu Asp Ala Asn Asn Cys Asp Tyr Glu Gln Lys

```
                    725                         730                         735
Glu Arg Phe Leu Leu Phe Cys Val His Gly Asp Ala Arg Gln Asp
                    740                         745                         750
Ser Leu Val Gln Trp Glu Met Glu Val Cys Lys Leu Pro Arg Leu
                    755                         760                         765
Ser Leu Asn Gly Val Arg Phe Lys Arg Ile Ser Gly Thr Ser Ile
                    770                         775                         780
Ala Phe Lys Asn Ile Ala Ser Lys Ile Ala Asn Glu Leu Lys Leu
                    785                         790                         795


<210> 5
<211> 656
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1242491CD1

<400> 5
Met Met Ser Trp Asn Leu Asn Lys Leu Gln Ser Phe Leu Leu Gly
  1                 5                          10                          15
Asp Gly Ser Phe Gly Ser Val Tyr Arg Ala Ala Tyr Glu Gly Glu
                    20                          25                          30
Glu Val Ala Val Lys Ile Phe Asn Lys His Thr Ser Leu Arg Leu
                    35                          40                          45
Leu Arg Gln Glu Leu Val Val Leu Cys His Leu His His Pro Ser
                    50                          55                          60
Leu Ile Ser Leu Leu Ala Ala Gly Ile Arg Pro Arg Met Leu Val
                    65                          70                          75
Met Glu Leu Ala Ser Lys Gly Ser Leu Asp Arg Leu Leu Gln Gln
                    80                          85                          90
Asp Lys Ala Ser Leu Thr Arg Thr Leu Gln His Arg Ile Ala Leu
                    95                          100                         105
His Val Ala Asp Gly Leu Arg Tyr Leu His Ser Ala Met Ile Ile
                    110                         115                         120
Tyr Arg Asp Leu Lys Pro His Asn Val Leu Leu Phe Thr Leu Tyr
                    125                         130                         135
Pro Asn Ala Ala Ile Ile Ala Lys Ile Ala Asp Tyr Gly Ile Ala
                    140                         145                         150
Gln Tyr Cys Cys Arg Met Gly Ile Lys Thr Ser Glu Gly Thr Pro
                    155                         160                         165
Gly Phe Arg Ala Pro Glu Val Ala Arg Gly Asn Val Ile Tyr Asn
                    170                         175                         180
Gln Gln Ala Asp Val Tyr Ser Phe Gly Leu Leu Leu Tyr Asp Ile
                    185                         190                         195
Leu Thr Thr Gly Gly Arg Ile Val Glu Gly Leu Lys Phe Pro Asn
                    200                         205                         210
Glu Phe Asp Glu Leu Glu Ile Gln Gly Lys Leu Pro Asp Pro Val
                    215                         220                         225
Lys Glu Tyr Gly Cys Ala Pro Trp Pro Met Val Glu Lys Leu Ile
                    230                         235                         240
Lys Gln Cys Leu Lys Glu Asn Pro Gln Glu Arg Pro Thr Ser Ala
                    245                         250                         255
Gln Val Phe Asp Ile Leu Asn Ser Ala Glu Leu Val Cys Leu Thr
                    260                         265                         270
Arg Arg Ile Leu Leu Pro Lys Asn Val Ile Val Glu Cys Met Val
                    275                         280                         285
Ala Thr His His Asn Ser Arg Asn Ala Ser Ile Trp Leu Gly Cys
                    290                         295                         300
Gly His Thr Asp Arg Gly Gln Leu Ser Phe Leu Asp Leu Asn Thr
                    305                         310                         315
Glu Gly Tyr Thr Ser Glu Glu Val Ala Asp Ser Arg Ile Leu Cys
                    320                         325                         330
Leu Ala Leu Val His Leu Pro Val Glu Lys Glu Ser Trp Ile Val
                    335                         340                         345
```

```
Ser Gly Thr Gln Ser Gly Thr Leu Leu Val Ile Asn Thr Glu Asp
                350                 355                 360
Gly Lys Lys Arg His Thr Leu Glu Lys Met Thr Asp Ser Val Thr
                365                 370                 375
Cys Leu Tyr Cys Asn Ser Phe Ser Lys Gln Ser Lys Gln Lys Asn
                380                 385                 390
Phe Leu Leu Val Gly Thr Ala Asp Gly Lys Leu Ala Ile Phe Glu
                395                 400                 405
Asp Lys Thr Val Lys Leu Lys Gly Ala Ala Pro Leu Lys Ile Leu
                410                 415                 420
Asn Ile Gly Asn Val Ser Thr Pro Leu Met Cys Leu Ser Glu Ser
                425                 430                 435
Thr Asn Ser Thr Glu Arg Asn Val Met Trp Gly Gly Cys Gly Thr
                440                 445                 450
Lys Ile Phe Ser Phe Ser Asn Asp Phe Thr Ile Gln Lys Leu Ile
                455                 460                 465
Glu Thr Arg Thr Ser Gln Leu Phe Ser Tyr Ala Ala Phe Ser Asp
                470                 475                 480
Ser Asn Ile Ile Thr Val Val Val Asp Thr Ala Leu Tyr Ile Ala
                485                 490                 495
Lys Gln Asn Ser Pro Val Val Glu Val Trp Asp Lys Lys Thr Glu
                500                 505                 510
Lys Leu Cys Gly Leu Ile Asp Cys Val His Phe Leu Arg Glu Val
                515                 520                 525
Thr Val Lys Glu Asn Lys Glu Ser Lys His Lys Met Ser Tyr Ser
                530                 535                 540
Gly Arg Val Lys Thr Leu Cys Leu Gln Lys Asn Thr Ala Leu Trp
                545                 550                 555
Ile Gly Thr Gly Gly Gly His Ile Leu Leu Leu Asp Leu Ser Thr
                560                 565                 570
Arg Arg Leu Ile Arg Val Ile Tyr Asn Phe Cys Asn Ser Val Arg
                575                 580                 585
Val Met Met Thr Ala Gln Leu Gly Ser Leu Lys Asn Val Met Leu
                590                 595                 600
Val Leu Gly Tyr Asn Arg Lys Asn Thr Glu Gly Thr Gln Lys Gln
                605                 610                 615
Lys Glu Ile Gln Ser Cys Leu Thr Val Trp Asp Ile Asn Leu Pro
                620                 625                 630
His Glu Val Gln Asn Leu Glu Lys His Ile Glu Val Arg Lys Glu
                635                 640                 645
Leu Ala Glu Lys Met Arg Arg Thr Ser Val Glu
                650                 655
```

<210> 6
<211> 596
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2634875CD1

<400> 6

```
Met Ala Thr Glu Asn Gly Ala Val Glu Leu Gly Ile Gln Asn Pro
  1               5                  10                  15
Ser Thr Asp Lys Ala Pro Lys Gly Pro Thr Gly Glu Arg Pro Leu
                 20                  25                  30
Ala Ala Gly Lys Asp Pro Gly Pro Pro Asp Pro Lys Lys Ala Pro
                 35                  40                  45
Asp Pro Pro Thr Leu Lys Lys Asp Ala Lys Ala Pro Ala Ser Glu
                 50                  55                  60
Lys Gly Asp Gly Thr Leu Ala Gln Pro Ser Thr Ser Ser Gln Gly
                 65                  70                  75
Pro Lys Gly Glu Gly Asp Arg Gly Gly Gly Pro Ala Glu Gly Ser
                 80                  85                  90
Ala Gly Pro Pro Ala Ala Leu Pro Gln Gln Thr Ala Thr Pro Glu
                 95                 100                 105
```

Thr Ser Val Lys Lys Pro Lys Ala Glu Gln Gly Ala Ser Gly Ser
110 115 120
Gln Asp Pro Gly Lys Pro Arg Val Gly Lys Lys Ala Ala Glu Gly
125 130 135
Gln Ala Ala Ala Arg Arg Gly Ser Pro Ala Phe Leu His Ser Pro
140 145 150
Ser Cys Pro Ala Ile Ile Ser Ser Ser Glu Lys Leu Leu Ala Lys
155 160 165
Lys Pro Pro Ser Glu Ala Ser Glu Leu Thr Phe Glu Gly Val Pro
170 175 180
Met Thr His Ser Pro Thr Asp Pro Arg Pro Ala Lys Ala Glu Glu
185 190 195
Gly Lys Asn Ile Leu Ala Glu Ser Gln Lys Glu Val Gly Glu Lys
200 205 210
Thr Pro Gly Gln Ala Gly Gln Ala Lys Met Gln Gly Asp Thr Ser
215 220 225
Arg Gly Ile Glu Phe Gln Ala Val Pro Ser Glu Lys Ser Glu Val
230 235 240
Gly Gln Ala Leu Cys Leu Thr Ala Arg Glu Glu Asp Cys Phe Gln
245 250 255
Ile Leu Asp Asp Cys Pro Pro Pro Pro Ala Pro Phe Pro His Arg
260 265 270
Met Val Glu Leu Arg Thr Gly Asn Val Ser Ser Glu Phe Ser Met
275 280 285
Asn Ser Lys Glu Ala Leu Gly Gly Gly Lys Phe Gly Ala Val Cys
290 295 300
Thr Cys Met Glu Lys Ala Thr Gly Leu Lys Leu Ala Ala Lys Val
305 310 315
Ile Lys Lys Gln Thr Pro Lys Asp Lys Glu Met Val Leu Leu Glu
320 325 330
Ile Glu Val Met Asn Gln Leu Asn His Arg Asn Leu Ile Gln Leu
335 340 345
Tyr Ala Ala Ile Glu Thr Pro His Glu Ile Val Leu Phe Met Glu
350 355 360
Tyr Ile Glu Gly Gly Glu Leu Phe Glu Arg Ile Val Asp Glu Asp
365 370 375
Tyr His Leu Thr Glu Val Asp Thr Met Val Phe Val Arg Gln Ile
380 385 390
Cys Asp Gly Ile Leu Phe Ser Val Leu Glu Arg Val Leu His Leu
395 400 405
Asp Leu Lys Pro Glu Asn Ile Leu Cys Val Asn Thr Thr Gly His
410 415 420
Leu Val Lys Ile Ile Asp Phe Gly Leu Ala Arg Arg Tyr Asn Pro
425 430 435
Asn Glu Lys Leu Lys Val Asn Phe Gly Thr Pro Glu Phe Leu Ser
440 445 450
Pro Glu Val Val Lys Gly Asp Gln Ile Ser Asp Lys Thr Asp Met
455 460 465
Trp Ser Met Gly Val Ile Thr Tyr Met Leu Leu Ser Gly Leu Ser
470 475 480
Pro Phe Leu Gly Asp Asp Asp Thr Glu Thr Leu Asn Asn Val Leu
485 490 495
Ser Gly Asn Trp Tyr Phe Asp Glu Glu Thr Phe Glu Ala Val Ser
500 505 510
Asp Glu Ala Lys Asp Phe Val Ser Asn Leu Ile Val Lys Asp Gln
515 520 525
Arg Ala Arg Met Asn Ala Ala Gln Cys Leu Ala His Pro Trp Leu
530 535 540
Asn Asn Leu Ala Glu Lys Ala Lys Arg Cys Asn Arg Arg Leu Lys
545 550 555
Ser Gln Ile Leu Leu Lys Lys Tyr Leu Met Lys Arg Arg Trp Lys
560 565 570
Lys Asn Phe Ile Ala Val Ser Ala Ala Asn Arg Phe Lys Lys Ile
575 580 585
Ser Ser Ser Gly Ala Leu Met Ala Leu Gly Val
590 595

<210> 7
<211> 497
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3951059CD1

<400> 7
```
Met Leu Lys Phe Lys Tyr Gly Ala Arg Asn Pro Leu Asp Ala Gly
 1               5                   10                  15
Ala Ala Glu Pro Ile Ala Ser Arg Ala Ser Arg Leu Asn Leu Phe
                20                  25                  30
Phe Gln Gly Lys Pro Pro Phe Met Thr Gln Gln Gln Met Ser Pro
                35                  40                  45
Leu Ser Arg Glu Gly Ile Leu Asp Ala Leu Phe Val Leu Phe Glu
                50                  55                  60
Glu Cys Ser Gln Pro Ala Leu Met Lys Ile Lys His Val Ser Asn
                65                  70                  75
Phe Val Arg Lys Tyr Ser Asp Thr Ile Ala Glu Leu Gln Glu Leu
                80                  85                  90
Gln Pro Ser Ala Lys Asp Phe Glu Val Arg Ser Leu Val Gly Cys
                95                  100                 105
Gly His Phe Ala Glu Val Gln Val Val Arg Glu Lys Ala Thr Gly
                110                 115                 120
Asp Ile Tyr Ala Met Lys Val Met Lys Lys Lys Ala Leu Leu Ala
                125                 130                 135
Gln Glu Gln Val Ser Phe Phe Glu Glu Glu Arg Asn Ile Leu Ser
                140                 145                 150
Arg Ser Thr Ser Pro Trp Ile Pro Gln Leu Gln Tyr Ala Phe Gln
                155                 160                 165
Asp Lys Asn His Leu Tyr Leu Val Met Glu Tyr Gln Pro Gly Gly
                170                 175                 180
Asp Leu Leu Ser Leu Leu Asn Arg Tyr Glu Asp Gln Leu Asp Glu
                185                 190                 195
Asn Leu Ile Gln Phe Tyr Leu Ala Glu Leu Ile Leu Ala Val His
                200                 205                 210
Ser Val His Leu Met Gly Tyr Val His Arg Asp Ile Lys Pro Glu
                215                 220                 225
Asn Ile Leu Val Asp Arg Thr Gly His Ile Lys Leu Val Asp Phe
                230                 235                 240
Gly Ser Ala Ala Lys Met Asn Ser Asn Lys Met Val Asn Ala Lys
                245                 250                 255
Leu Pro Ile Gly Thr Pro Asp Tyr Met Ala Pro Glu Val Leu Thr
                260                 265                 270
Val Met Asn Gly Asp Gly Lys Gly Thr Tyr Arg Leu Asp Cys Asp
                275                 280                 285
Trp Trp Ser Val Gly Val Ile Ala Tyr Glu Met Ile Tyr Gly Arg
                290                 295                 300
Ser Pro Phe Ala Glu Gly Thr Ser Ala Arg Thr Phe Asn Asn Ile
                305                 310                 315
Met Asn Phe Gln Arg Phe Leu Lys Phe Pro Asp Asp Pro Lys Val
                320                 325                 330
Ser Ser Asp Phe Leu Asp Leu Ile Gln Ser Leu Leu Cys Gly Gln
                335                 340                 345
Lys Glu Arg Leu Lys Phe Glu Gly Leu Cys Cys His Pro Phe Phe
                350                 355                 360
Ser Lys Ile Asp Trp Asn Asn Ile Arg Asn Ser Pro Pro Pro Phe
                365                 370                 375
Val Pro Thr Leu Lys Ser Asp Asp Asp Thr Ser Asn Phe Asp Glu
                380                 385                 390
Pro Glu Lys Asn Ser Trp Val Ser Ser Ser Pro Cys Gln Leu Ser
                395                 400                 405
Pro Ser Gly Phe Ser Gly Glu Glu Leu Pro Phe Val Gly Phe Ser
                410                 415                 420
Tyr Ser Lys Ala Leu Gly Ile Leu Gly Arg Ser Glu Ser Val Val
```

```
                425                 430                 435
Ser Gly Leu Asp Ser Pro Ala Lys Thr Ser Ser Met Glu Lys Lys
                440                 445                 450
Leu Leu Ile Lys Ser Lys Glu Leu Gln Asp Ser Gln Asp Lys Cys
                455                 460                 465
His Lys Val Phe Ile Ser Ala Ala Gly Leu Leu Pro Cys Ser Arg
                470                 475                 480
Ile Leu Pro Ser Val Tyr Ala Lys Gly Ser Ala Arg Gly Arg Cys
                485                 490                 495
Trp Leu


<210> 8
<211> 1171
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7395890CD1

<400> 8
Met Ala Pro Val Tyr Glu Gly Met Ala Ser His Val Gln Val Phe
  1               5                  10                  15
Ser Pro His Thr Leu Gln Ser Ser Ala Phe Cys Ser Val Lys Lys
                 20                  25                  30
Leu Lys Ile Glu Pro Ser Ser Asn Trp Asp Met Thr Gly Tyr Gly
                 35                  40                  45
Ser His Ser Lys Val Tyr Ser Gln Ser Lys Asn Ile Pro Leu Ser
                 50                  55                  60
Gln Pro Ala Thr Thr Thr Val Ser Thr Ser Leu Pro Val Pro Asn
                 65                  70                  75
Pro Ser Leu Pro Tyr Glu Gln Thr Ile Val Phe Pro Gly Ser Thr
                 80                  85                  90
Gly His Ile Val Val Thr Ser Ala Ser Ser Thr Ser Val Thr Gly
                 95                 100                 105
Gln Val Leu Gly Gly Pro His Asn Leu Met Arg Arg Ser Thr Val
                110                 115                 120
Ser Leu Leu Asp Thr Tyr Gln Lys Cys Gly Leu Lys Arg Lys Ser
                125                 130                 135
Glu Glu Ile Glu Asn Thr Ser Ser Val Gln Ile Ile Glu Glu His
                140                 145                 150
Pro Pro Met Ile Gln Asn Asn Ala Ser Gly Ala Thr Val Ala Thr
                155                 160                 165
Ala Thr Thr Ser Thr Ala Thr Ser Lys Asn Ser Gly Ser Asn Ser
                170                 175                 180
Glu Gly Asp Tyr Gln Leu Val Gln His Glu Val Leu Cys Ser Met
                185                 190                 195
Thr Asn Thr Tyr Glu Val Leu Glu Phe Leu Gly Arg Gly Thr Phe
                200                 205                 210
Gly Gln Val Val Lys Cys Trp Lys Arg Gly Thr Asn Glu Ile Val
                215                 220                 225
Ala Ile Lys Ile Leu Lys Asn His Pro Ser Tyr Ala Arg Gln Gly
                230                 235                 240
Gln Ile Glu Val Ser Ile Leu Ala Arg Leu Ser Thr Glu Ser Ala
                245                 250                 255
Asp Asp Tyr Asn Phe Val Arg Ala Tyr Glu Cys Phe Gln His Lys
                260                 265                 270
Asn His Thr Cys Leu Val Phe Glu Met Leu Glu Gln Asn Leu Tyr
                275                 280                 285
Asp Phe Leu Lys Gln Asn Lys Phe Ser Pro Leu Pro Leu Lys Tyr
                290                 295                 300
Ile Arg Pro Val Leu Gln Gln Val Ala Thr Ala Leu Met Lys Leu
                305                 310                 315
Lys Ser Leu Gly Leu Ile His Ala Asp Leu Lys Pro Glu Asn Ile
                320                 325                 330
Met Leu Val Asp Pro Ser Arg Gln Pro Tyr Arg Val Lys Val Ile
```

```
                    335                           340                           345
Asp Phe Gly Ser Ala Ser His Val Ser Lys Ala Val Cys Ser Thr
                    350                           355                           360
Tyr Leu Gln Ser Arg Tyr Tyr Arg Ala Pro Glu Ile Ile Leu Gly
                    365                           370                           375
Leu Pro Phe Cys Glu Ala Ile Asp Met Trp Ser Leu Gly Cys Val
                    380                           385                           390
Ile Ala Glu Leu Phe Leu Gly Trp Pro Leu Tyr Pro Gly Ala Ser
                    395                           400                           405
Glu Tyr Asp Gln Ile Arg Tyr Ile Ser Gln Thr Gln Gly Leu Pro
                    410                           415                           420
Ala Glu Tyr Leu Leu Ser Ala Gly Thr Lys Thr Thr Arg Phe Phe
                    425                           430                           435
Asn Arg Asp Thr Asp Ser Pro Tyr Pro Leu Trp Arg Leu Lys Thr
                    440                           445                           450
Pro Asp Asp His Glu Ala Glu Thr Gly Ile Lys Ser Lys Glu Ala
                    455                           460                           465
Arg Lys Tyr Ile Phe Asn Cys Leu Asp Asp Met Ala Gln Val Asn
                    470                           475                           480
Met Thr Thr Asp Leu Glu Gly Ser Asp Met Leu Val Glu Lys Ala
                    485                           490                           495
Asp Arg Arg Glu Phe Ile Asp Leu Leu Lys Lys Met Leu Thr Ile
                    500                           505                           510
Asp Ala Asp Lys Arg Ile Thr Pro Ile Glu Thr Leu Asn His Pro
                    515                           520                           525
Phe Val Thr Met Thr His Leu Leu Asp Phe Pro His Ser Thr His
                    530                           535                           540
Val Lys Ser Cys Phe Gln Asn Met Glu Ile Cys Lys Arg Arg Val
                    545                           550                           555
Asn Met Tyr Asp Thr Val Asn Gln Ser Lys Thr Pro Phe Ile Thr
                    560                           565                           570
His Val Ala Pro Ser Thr Ser Thr Asn Leu Thr Met Thr Phe Asn
                    575                           580                           585
Asn Gln Leu Thr Thr Val His Asn Gln Pro Ser Ala Ala Ser Met
                    590                           595                           600
Ala Ala Val Ala Gln Arg Ser Met Pro Leu Gln Thr Gly Thr Ala
                    605                           610                           615
Gln Ile Cys Ala Arg Pro Asp Pro Phe Gln Gln Ala Leu Ile Val
                    620                           625                           630
Cys Pro Pro Gly Phe Gln Gly Leu Gln Ala Ser Pro Ser Lys His
                    635                           640                           645
Ala Gly Tyr Ser Val Arg Met Glu Asn Ala Val Pro Ile Val Thr
                    650                           655                           660
Gln Ala Pro Gly Ala Gln Pro Leu Gln Ile Gln Pro Gly Leu Leu
                    665                           670                           675
Ala Gln Gln Ala Trp Pro Ser Gly Thr Gln Gln Ile Leu Leu Pro
                    680                           685                           690
Pro Ala Trp Gln Gln Leu Thr Gly Val Ala Thr His Thr Ser Val
                    695                           700                           705
Gln His Ala Thr Val Ile Pro Glu Thr Met Ala Gly Thr Gln Gln
                    710                           715                           720
Leu Ala Asp Trp Arg Asn Thr His Ala His Gly Ser His Tyr Asn
                    725                           730                           735
Pro Ile Met Gln Gln Pro Ala Leu Leu Thr Gly His Val Thr Leu
                    740                           745                           750
Pro Ala Ala Gln Pro Leu Asn Val Gly Val Ala His Val Met Arg
                    755                           760                           765
Gln Gln Pro Thr Ser Thr Thr Ser Ser Arg Lys Ser Lys Gln His
                    770                           775                           780
Gln Ser Ser Val Arg Asn Val Ser Thr Cys Glu Val Ser Ser Ser
                    785                           790                           795
Gln Ala Ile Ser Ser Pro Gln Arg Ser Lys Arg Val Lys Glu Asn
                    800                           805                           810
Thr Pro Pro Arg Cys Ala Met Val His Ser Ser Pro Ala Cys Ser
                    815                           820                           825
Thr Ser Val Thr Cys Gly Trp Gly Asp Val Ala Ser Ser Thr Thr
                    830                           835                           840
```

```
Arg Glu Arg Gln Arg Gln Thr Ile Val Ile Pro Asp Thr Pro Ser
                845                 850                 855
Pro Thr Val Ser Val Ile Thr Ile Ser Ser Asp Thr Asp Glu Glu
                860                 865                 870
Glu Glu Gln Lys His Ala Pro Thr Ser Thr Val Ser Lys Gln Arg
                875                 880                 885
Lys Asn Val Ile Ser Cys Val Thr Val His Asp Ser Pro Tyr Ser
                890                 895                 900
Asp Ser Ser Ser Asn Thr Ser Pro Tyr Ser Val Gln Gln Arg Ala
                905                 910                 915
Gly His Asn Asn Ala Asn Ala Phe Asp Thr Lys Gly Ser Leu Glu
                920                 925                 930
Asn His Cys Thr Gly Asn Pro Arg Thr Ile Ile Val Pro Pro Leu
                935                 940                 945
Lys Thr Gln Ala Ser Glu Val Leu Val Glu Cys Asp Ser Leu Val
                950                 955                 960
Pro Val Asn Thr Ser His His Ser Ser Ser Tyr Lys Ser Lys Ser
                965                 970                 975
Ser Ser Asn Val Thr Ser Thr Ser Gly His Ser Ser Gly Ser Ser
                980                 985                 990
Ser Gly Ala Ile Thr Tyr Arg Gln Gln Arg Pro Gly Pro His Phe
                995                1000                1005
Gln Gln Gln Gln Pro Leu Asn Leu Ser Gln Ala Gln Gln His Ile
               1010                1015                1020
Thr Thr Asp Arg Thr Gly Ser His Arg Arg Gln Gln Ala Tyr Ile
               1025                1030                1035
Thr Pro Thr Met Ala Gln Ala Pro Tyr Ser Phe Pro His Asn Ser
               1040                1045                1050
Pro Ser His Gly Thr Val His Pro His Leu Ala Ala Ala Ala Ala
               1055                1060                1065
Ala Ala His Leu Pro Thr Gln Pro His Leu Tyr Thr Tyr Thr Ala
               1070                1075                1080
Pro Ala Ala Leu Gly Ser Thr Gly Thr Val Ala His Leu Val Ala
               1085                1090                1095
Ser Gln Gly Ser Ala Arg His Thr Val Gln His Thr Ala Tyr Pro
               1100                1105                1110
Ala Ser Ile Val His Gln Val Pro Val Ser Met Gly Pro Arg Val
               1115                1120                1125
Leu Pro Ser Pro Thr Ile His Pro Ser Gln Tyr Pro Ala Gln Phe
               1130                1135                1140
Ala His Gln Thr Tyr Ile Ser Ala Ser Pro Ala Ser Thr Val Tyr
               1145                1150                1155
Thr Gly Tyr Pro Leu Ser Pro Ala Lys Val Asn Gln Tyr Pro Tyr
               1160                1165                1170
Ile
```

<210> 9
<211> 470
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475546CD1

<400> 9

```
Met Ala Gly Pro Gly Trp Gly Pro Pro Arg Leu Asp Gly Phe Ile
  1               5                  10                  15
Leu Thr Glu Arg Leu Gly Ser Gly Thr Tyr Ala Thr Val Tyr Lys
                 20                  25                  30
Ala Tyr Ala Lys Lys Asp Thr Arg Glu Val Val Ala Ile Lys Cys
                 35                  40                  45
Val Ala Lys Lys Ser Leu Asn Lys Ala Ser Val Glu Asn Leu Leu
                 50                  55                  60
Thr Glu Ile Glu Ile Leu Lys Gly Ile Arg His Pro His Ile Val
                 65                  70                  75
```

```
Gln Leu Lys Asp Phe Gln Trp Asp Ser Asp Asn Ile Tyr Leu Ile
                80                  85                  90
Met Glu Phe Cys Ala Gly Gly Asp Leu Ser Arg Phe Ile His Thr
                95                  100                 105
Arg Arg Ile Leu Pro Glu Lys Val Ala Arg Val Phe Met Gln Gln
                110                 115                 120
Leu Ala Ser Ala Leu Gln Phe Leu His Glu Arg Asn Ile Ser His
                125                 130                 135
Leu Asp Leu Lys Pro Gln Asn Ile Leu Leu Ser Ser Leu Glu Lys
                140                 145                 150
Pro His Leu Lys Leu Ala Asp Phe Gly Phe Ala Gln His Met Ser
                155                 160                 165
Pro Trp Asp Glu Lys His Val Leu Arg Gly Ser Pro Leu Tyr Met
                170                 175                 180
Ala Pro Glu Met Val Cys Gln Arg Gln Tyr Asp Ala Arg Val Asp
                185                 190                 195
Leu Trp Ser Met Gly Val Ile Leu Tyr Glu Ala Leu Phe Gly Gln
                200                 205                 210
Pro Pro Phe Ala Ser Arg Ser Phe Ser Glu Leu Glu Glu Lys Ile
                215                 220                 225
Arg Ser Asn Arg Val Ile Glu Leu Pro Leu Arg Pro Leu Leu Ser
                230                 235                 240
Arg Asp Cys Arg Asp Leu Leu Gln Arg Leu Leu Glu Arg Asp Pro
                245                 250                 255
Ser Arg Arg Ile Ser Phe Gln Asp Phe Phe Ala His Pro Trp Val
                260                 265                 270
Asp Leu Glu His Met Pro Ser Gly Glu Ser Leu Gly Arg Ala Thr
                275                 280                 285
Ala Leu Val Val Gln Ala Val Lys Lys Asp Gln Glu Gly Asp Ser
                290                 295                 300
Ala Ala Ala Leu Ser Leu Tyr Cys Lys Ala Leu Asp Phe Phe Val
                305                 310                 315
Pro Ala Leu His Tyr Glu Val Asp Ala Gln Arg Lys Glu Ala Ile
                320                 325                 330
Lys Ala Lys Val Gly Gln Tyr Val Ser Arg Ala Glu Glu Leu Lys
                335                 340                 345
Ala Ile Val Ser Ser Ser Asn Gln Ala Leu Leu Arg Gln Gly Thr
                350                 355                 360
Ser Ala Arg Asp Leu Leu Arg Glu Met Ala Arg Asp Lys Pro Arg
                365                 370                 375
Leu Leu Ala Ala Leu Glu Val Ala Ser Ala Ala Met Ala Lys Glu
                380                 385                 390
Glu Ala Ala Gly Gly Glu Gln Asp Ala Leu Asp Leu Tyr Gln His
                395                 400                 405
Ser Leu Gly Glu Leu Leu Leu Leu Leu Ala Ala Glu Pro Pro Gly
                410                 415                 420
Arg Arg Arg Glu Leu Leu His Thr Glu Val Gln Asn Leu Met Ala
                425                 430                 435
Arg Ala Glu Tyr Leu Lys Glu Gln Met Arg Glu Ser Arg Trp Glu
                440                 445                 450
Ala Asp Thr Leu Asp Lys Glu Gly Leu Ser Glu Ser Val Arg Ser
                455                 460                 465
Ser Cys Thr Leu Gln
                470
```

<210> 10
<211> 422
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477076CD1

<400> 10

```
Met Asp His Pro Ser Arg Glu Lys Asp Glu Arg Gln Arg Thr Thr
  1               5                  10                  15
```

```
Lys Pro Met Ala Gln Arg Ser Ala His Cys Ser Arg Pro Ser Gly
                20                  25                  30
Ser Ser Ser Ser Ser Gly Val Leu Met Val Gly Pro Asn Phe Arg
                35                  40                  45
Val Gly Lys Lys Ile Gly Cys Gly Asn Phe Gly Glu Leu Arg Leu
                50                  55                  60
Gly Lys Asn Leu Tyr Thr Asn Glu Tyr Val Ala Ile Lys Leu Glu
                65                  70                  75
Pro Ile Lys Ser Arg Ala Pro Gln Leu His Leu Glu Tyr Arg Phe
                80                  85                  90
Tyr Lys Gln Leu Gly Ser Ala Gly Glu Gly Leu Pro Gln Val Tyr
                95                  100                 105
Tyr Phe Gly Pro Cys Gly Lys Tyr Asn Ala Met Val Leu Glu Leu
                110                 115                 120
Leu Gly Pro Ser Leu Glu Asp Leu Phe Asp Leu Cys Asp Arg Thr
                125                 130                 135
Phe Thr Leu Lys Thr Val Leu Met Ile Ala Ile Gln Leu Leu Ser
                140                 145                 150
Arg Met Glu Tyr Val His Ser Lys Asn Leu Ile Tyr Arg Asp Val
                155                 160                 165
Lys Pro Glu Asn Phe Leu Ile Gly Arg Gln Gly Asn Lys Lys Glu
                170                 175                 180
His Val Ile His Ile Ile Asp Phe Gly Leu Ala Lys Glu Tyr Ile
                185                 190                 195
Asp Pro Glu Thr Lys Lys His Ile Pro Tyr Arg Glu His Lys Ser
                200                 205                 210
Leu Thr Gly Thr Ala Arg Tyr Met Ser Ile Asn Thr His Leu Gly
                215                 220                 225
Lys Glu Gln Ser Arg Arg Asp Asp Leu Glu Ala Leu Gly His Met
                230                 235                 240
Phe Met Tyr Phe Leu Arg Gly Ser Leu Pro Trp Gln Gly Leu Lys
                245                 250                 255
Ala Asp Thr Leu Lys Glu Arg Tyr Gln Lys Ile Gly Asp Thr Lys
                260                 265                 270
Arg Asn Thr Pro Ile Glu Ala Leu Cys Glu Asn Phe Pro Glu Glu
                275                 280                 285
Met Ala Thr Tyr Leu Arg Tyr Val Arg Arg Leu Asp Phe Phe Glu
                290                 295                 300
Lys Pro Asp Tyr Glu Tyr Leu Arg Thr Leu Phe Thr Asp Leu Phe
                305                 310                 315
Glu Lys Lys Gly Tyr Thr Phe Asp Tyr Ala Tyr Asp Trp Val Gly
                320                 325                 330
Arg Pro Ile Pro Thr Pro Val Gly Ser Val His Val Asp Ser Gly
                335                 340                 345
Ala Ser Ala Ile Thr Arg Glu Ser His Thr His Arg Asp Arg Pro
                350                 355                 360
Ser Gln Gln Gln Pro Leu Arg Asn Gln Val Val Ser Ser Thr Asn
                365                 370                 375
Gly Glu Leu Asn Val Asp Asp Pro Thr Gly Ala His Ser Asn Ala
                380                 385                 390
Pro Ile Thr Ala His Ala Glu Val Glu Val Val Glu Glu Ala Lys
                395                 400                 405
Cys Cys Cys Phe Phe Lys Arg Lys Arg Lys Lys Thr Ala Gln Arg
                410                 415                 420
His Lys
```

<210> 11
<211> 240
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1874092CD1

<400> 11

Met Pro Val Ser Lys Cys Pro Lys Lys Ser Glu Ser Leu Trp Lys
1 5 10 15
Gly Trp Asp Arg Lys Ala Gln Arg Asn Gly Leu Arg Ser Gln Val
20 25 30
Tyr Ala Val Asn Gly Asp Tyr Tyr Val Gly Glu Trp Lys Asp Asn
35 40 45
Val Lys His Gly Lys Gly Thr Gln Val Trp Lys Lys Lys Gly Ala
50 55 60
Ile Tyr Glu Gly Asp Trp Lys Phe Gly Lys Arg Asp Gly Tyr Gly
65 70 75
Thr Leu Ser Leu Pro Asp Gln Gln Thr Gly Lys Cys Arg Arg Val
80 85 90
Tyr Ser Gly Trp Trp Lys Gly Asp Lys Lys Ser Gly Tyr Gly Ile
95 100 105
Gln Phe Phe Gly Pro Lys Glu Tyr Tyr Glu Gly Asp Trp Cys Gly
110 115 120
Ser Gln Arg Ser Gly Trp Gly Arg Met Tyr Tyr Ser Asn Gly Asp
125 130 135
Ile Tyr Glu Gly Gln Trp Glu Asn Asp Lys Pro Asn Gly Glu Gly
140 145 150
Met Leu Arg Leu Lys Asn Gly Asn Arg Tyr Glu Gly Cys Trp Glu
155 160 165
Arg Gly Met Lys Asn Gly Ala Gly Arg Phe Phe His Leu Asp His
170 175 180
Gly Gln Leu Phe Glu Gly Phe Trp Val Asp Asn Met Ala Lys Cys
185 190 195
Gly Thr Met Ile Asp Phe Gly Arg Asp Glu Ala Pro Glu Pro Thr
200 205 210
Gln Phe Pro Ile Pro Glu Val Lys Ile Leu Asp Pro Asp Gly Val
215 220 225
Leu Ala Glu Ala Leu Ala Met Phe Arg Lys Thr Glu Glu Gly Asp
230 235 240

<210> 12
<211> 594
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4841542CD1

<400> 12
Met Lys Lys Gln Ala Val Lys Arg His His His Lys His Asn Leu
1 5 10 15
Arg His Arg Tyr Glu Phe Leu Glu Thr Leu Gly Lys Gly Thr Tyr
20 25 30
Gly Lys Val Lys Lys Ala Arg Glu Ser Ser Gly Arg Leu Val Ala
35 40 45
Ile Lys Ser Ile Arg Lys Asp Lys Ile Lys Asp Glu Gln Asp Leu
50 55 60
Met His Ile Arg Arg Glu Ile Glu Ile Met Ser Ser Leu Asn His
65 70 75
Pro His Ile Ile Ala Ile His Glu Val Phe Glu Asn Ser Ser Lys
80 85 90
Ile Val Ile Val Met Glu Tyr Ala Ser Arg Gly Asp Leu Tyr Asp
95 100 105
Tyr Ile Ser Glu Arg Gln Gln Leu Ser Glu Arg Glu Ala Arg His
110 115 120
Phe Phe Arg Gln Ile Val Ser Ala Val His Tyr Cys His Gln Asn
125 130 135
Arg Val Val His Arg Asp Leu Lys Leu Glu Asn Ile Leu Leu Gly
140 145 150
Ala Asn Gly Asn Ile Lys Ile Ala Asp Phe Gly Leu Ser Asn Leu
155 160 165
Tyr His Gln Gly Lys Phe Leu Gln Thr Phe Cys Gly Ser Pro Leu

```
                170                     175                     180
Tyr Ala Ser Pro Glu Ile Val Asn Gly Lys Pro Tyr Thr Gly Pro
                185                     190                     195
Glu Val Asp Ser Trp Ser Leu Gly Val Leu Leu Tyr Ile Leu Val
                200                     205                     210
His Gly Thr Met Pro Phe Asp Gly His Asp His Lys Ile Leu Val
                215                     220                     225
Lys Gln Ile Ser Asn Gly Ala Tyr Arg Glu Pro Pro Lys Pro Ser
                230                     235                     240
Asp Ala Cys Gly Leu Ile Arg Trp Leu Leu Met Val Asn Pro Thr
                245                     250                     255
Arg Arg Ala Thr Leu Glu Asp Val Ala Ser His Trp Trp Val Asn
                260                     265                     270
Trp Gly Tyr Ala Thr Arg Val Gly Glu Gln Glu Ala Pro His Glu
                275                     280                     285
Gly Gly His Pro Gly Ser Asp Ser Ala Arg Ala Ser Met Ala Asp
                290                     295                     300
Trp Leu Arg Arg Ser Ser Arg Pro Leu Leu Glu Asn Gly Ala Lys
                305                     310                     315
Val Cys Ser Phe Phe Lys Gln His Ala Pro Gly Gly Gly Ser Thr
                320                     325                     330
Thr Pro Gly Leu Glu Arg Gln His Ser Leu Lys Lys Ser Arg Lys
                335                     340                     345
Glu Asn Asp Met Ala Gln Ser Leu His Ser Asp Thr Ala Asp Asp
                350                     355                     360
Thr Ala His Arg Pro Gly Lys Ser Asn Leu Lys Leu Pro Lys Gly
                365                     370                     375
Ile Leu Lys Lys Lys Val Ser Ala Ser Ala Glu Gly Val Gln Glu
                380                     385                     390
Asp Pro Pro Glu Leu Ser Pro Ile Pro Ala Ser Pro Gly Gln Ala
                395                     400                     405
Ala Pro Leu Leu Pro Lys Lys Gly Ile Leu Lys Lys Pro Arg Gln
                410                     415                     420
Arg Glu Ser Gly Tyr Tyr Ser Ser Pro Glu Pro Ser Glu Ser Gly
                425                     430                     435
Glu Leu Leu Asp Ala Gly Asp Val Phe Val Ser Gly Asp Pro Lys
                440                     445                     450
Glu Gln Lys Pro Pro Gln Ala Ser Gly Leu Leu Leu His Arg Lys
                455                     460                     465
Gly Ile Leu Lys Leu Asn Gly Lys Phe Ser Gln Thr Ala Leu Glu
                470                     475                     480
Leu Ala Ala Pro Thr Thr Phe Gly Ser Leu Asp Glu Leu Ala Pro
                485                     490                     495
Pro Arg Pro Leu Ala Arg Ala Ser Arg Pro Ser Gly Ala Val Ser
                500                     505                     510
Glu Asp Ser Ile Leu Ser Ser Glu Ser Phe Asp Gln Leu Asp Leu
                515                     520                     525
Pro Glu Arg Leu Pro Glu Pro Pro Leu Arg Gly Cys Val Ser Val
                530                     535                     540
Asp Asn Leu Thr Gly Leu Glu Glu Pro Pro Ser Glu Gly Pro Gly
                545                     550                     555
Ser Cys Leu Arg Arg Trp Arg Gln Asp Pro Leu Gly Asp Ser Cys
                560                     565                     570
Phe Ser Leu Thr Asp Cys Gln Glu Val Thr Ala Thr Tyr Arg Gln
                575                     580                     585
Ala Leu Arg Val Cys Ser Lys Leu Thr
                590
```

<210> 13
<211> 473
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472695CD1

<400> 13
Met Ser Gln Thr Ser Ser Ile Gly Ser Ala Glu Ser Leu Ile Ser
1 5 10 15
Leu Glu Arg Lys Lys Glu Lys Asn Ile Asn Arg Asp Ile Thr Ser
20 25 30
Arg Lys Asp Leu Pro Ser Arg Thr Ser Asn Val Glu Arg Lys Ala
35 40 45
Ser Gln Gln Gln Trp Gly Arg Gly Asn Phe Thr Glu Gly Lys Val
50 55 60
Pro His Ile Arg Ile Glu Asn Gly Ala Ala Ile Glu Glu Ile Tyr
65 70 75
Thr Phe Gly Arg Ile Leu Gly Lys Gly Ser Phe Gly Ile Val Ile
80 85 90
Glu Ala Thr Asp Lys Glu Thr Glu Thr Lys Trp Ala Ile Lys Lys
95 100 105
Val Asn Lys Glu Lys Ala Gly Ser Ser Ala Val Lys Leu Leu Glu
110 115 120
Arg Glu Val Asn Ile Leu Lys Ser Val Lys His Glu His Ile Ile
125 130 135
His Leu Glu Gln Val Phe Glu Thr Pro Lys Lys Met Tyr Leu Val
140 145 150
Met Glu Leu Cys Glu Asp Gly Glu Leu Lys Glu Ile Leu Asp Arg
155 160 165
Lys Gly His Phe Ser Glu Asn Glu Thr Arg Trp Ile Ile Gln Ser
170 175 180
Leu Ala Ser Ala Ile Ala Tyr Leu His Asn Asn Asp Ile Val His
185 190 195
Arg Asp Leu Lys Leu Glu Asn Ile Met Val Lys Ser Ser Leu Ile
200 205 210
Asp Asp Asn Asn Glu Ile Asn Leu Asn Ile Lys Val Thr Asp Phe
215 220 225
Gly Leu Ala Val Lys Lys Gln Ser Arg Ser Glu Ala Met Leu Gln
230 235 240
Ala Thr Cys Gly Thr Pro Ile Tyr Met Ala Pro Glu Val Ile Ser
245 250 255
Ala His Asp Tyr Ser Gln Gln Cys Asp Ile Trp Ser Ile Gly Val
260 265 270
Val Met Tyr Met Leu Leu Arg Gly Glu Pro Pro Phe Leu Ala Ser
275 280 285
Ser Glu Glu Lys Leu Phe Glu Leu Ile Arg Lys Gly Glu Leu His
290 295 300
Phe Glu Asn Ala Val Trp Asn Ser Ile Ser Asp Cys Ala Lys Ser
305 310 315
Val Leu Lys Gln Leu Met Lys Val Asp Pro Ala His Arg Ile Thr
320 325 330
Ala Lys Glu Leu Leu Asp Asn Gln Trp Leu Thr Gly Asn Lys Leu
335 340 345
Ser Ser Val Arg Pro Thr Asn Val Leu Glu Met Met Lys Glu Trp
350 355 360
Lys Asn Asn Pro Glu Ser Val Glu Glu Asn Thr Thr Glu Glu Lys
365 370 375
Asn Lys Pro Ser Thr Glu Glu Lys Leu Lys Ser Tyr Gln Pro Trp
380 385 390
Gly Asn Val Pro Asp Ala Asn Tyr Thr Ser Asp Glu Glu Glu Glu
395 400 405
Lys Gln Ser Thr Ala Tyr Glu Lys Gln Phe Pro Ala Thr Ser Lys
410 415 420
Asp Asn Phe Asp Met Cys Ser Ser Ser Phe Thr Ser Ser Lys Leu
425 430 435
Leu Pro Ala Glu Ile Lys Gly Glu Met Glu Lys Thr Pro Val Thr
440 445 450
Pro Ser Gln Gly Thr Ala Thr Lys Tyr Pro Ala Lys Ser Gly Ala
455 460 465
Leu Ser Arg Thr Lys Lys Lys Leu
470

<210> 14

<211> 947
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477966CD1

<400> 14
```
Met Met Ser Asp Thr Ser Thr Phe Pro Asn His Pro Ser Ser Pro
 1               5                  10                  15
Ala Ala Ser Pro Ser Gly Gly Arg Gly Val Met Ala Ser Pro Ala
                20                  25                  30
Trp Asp Arg Ser Lys Gly Trp Ser Gln Thr Pro Gln Arg Ala Asp
                35                  40                  45
Phe Val Ser Thr Pro Leu Gln Val His Thr Leu Arg Pro Glu Asn
                50                  55                  60
Leu Leu Leu Val Ser Thr Leu Asp Gly Ser Leu His Ala Leu Ser
                65                  70                  75
Lys Gln Thr Gly Asp Leu Lys Trp Thr Leu Arg Asp Asp Pro Val
                80                  85                  90
Ile Glu Gly Pro Met Tyr Val Thr Glu Met Ala Phe Leu Ser Asp
                95                 100                 105
Pro Ala Asp Gly Ser Leu Tyr Ile Leu Gly Thr Gln Lys Gln Gln
               110                 115                 120
Gly Leu Met Lys Leu Pro Phe Thr Ile Pro Glu Leu Val His Ala
               125                 130                 135
Ser Pro Cys Arg Ser Ser Asp Gly Val Phe Tyr Thr Gly Arg Lys
               140                 145                 150
Gln Asp Ala Trp Phe Val Val Asp Pro Glu Ser Gly Glu Thr Gln
               155                 160                 165
Met Thr Leu Thr Thr Glu Gly Pro Ser Thr Pro Arg Leu Tyr Ile
               170                 175                 180
Gly Arg Thr Gln Tyr Thr Val Thr Met His Asp Pro Arg Ala Pro
               185                 190                 195
Ala Leu Arg Trp Asn Thr Thr Tyr Arg Arg Tyr Ser Ala Pro Pro
               200                 205                 210
Met Asp Gly Ser Pro Gly Lys Tyr Met Ser His Leu Ala Ser Cys
               215                 220                 225
Gly Met Gly Leu Leu Leu Thr Val Asp Pro Gly Ser Gly Thr Val
               230                 235                 240
Leu Trp Thr Gln Asp Leu Gly Val Pro Val Met Gly Val Tyr Thr
               245                 250                 255
Trp His Gln Asp Gly Leu Arg Gln Leu Pro His Leu Thr Leu Ala
               260                 265                 270
Arg Asp Thr Leu His Phe Leu Ala Leu Arg Trp Gly His Ile Arg
               275                 280                 285
Leu Pro Ala Ser Gly Pro Arg Asp Thr Ala Thr Leu Phe Ser Thr
               290                 295                 300
Leu Asp Thr Gln Leu Leu Met Thr Leu Tyr Val Gly Lys Asp Glu
               305                 310                 315
Thr Gly Phe Tyr Val Ser Lys Ala Leu Val His Thr Gly Val Ala
               320                 325                 330
Leu Val Pro Arg Gly Leu Thr Leu Ala Pro Ala Asp Gly Pro Thr
               335                 340                 345
Thr Asp Glu Val Thr Leu Gln Val Ser Gly Glu Arg Glu Gly Ser
               350                 355                 360
Pro Ser Thr Ala Val Arg Tyr Pro Ser Gly Ser Val Ala Leu Pro
               365                 370                 375
Ser Gln Trp Leu Leu Ile Gly His His Glu Leu Pro Pro Val Leu
               380                 385                 390
His Thr Thr Met Leu Arg Val His Pro Thr Leu Gly Ser Gly Thr
               395                 400                 405
Ala Glu Thr Arg Pro Pro Glu Asn Thr Gln Ala Pro Ala Phe Phe
               410                 415                 420
Leu Glu Leu Leu Ser Leu Ser Arg Glu Lys Leu Trp Asp Ser Glu
               425                 430                 435
```

Leu His Pro Glu Glu Lys Thr Pro Asp Ser Tyr Leu Gly Leu Gly
440 445 450
Pro Gln Asp Leu Leu Ala Ala Ser Leu Thr Ala Val Leu Leu Gly
455 460 465
Gly Trp Ile Leu Phe Val Met Arg Gln Gln Gln Pro Gln Val Val
470 475 480
Glu Lys Gln Gln Glu Thr Pro Leu Ala Pro Ala Asp Phe Ala His
485 490 495
Ile Ser Gln Asp Ala Gln Ser Leu His Ser Gly Ala Ser Arg Arg
500 505 510
Ser Gln Lys Arg Leu Gln Ser Pro Ser Lys Gln Ala Gln Pro Leu
515 520 525
Asp Asp Pro Glu Ala Glu Gln Leu Thr Val Val Gly Lys Ile Ser
530 535 540
Phe Asn Pro Lys Asp Val Leu Gly Arg Gly Ala Gly Gly Thr Phe
545 550 555
Val Phe Arg Gly Gln Phe Glu Gly Arg Ala Val Ala Val Lys Arg
560 565 570
Leu Leu Arg Glu Cys Phe Gly Leu Val Arg Arg Glu Val Gln Leu
575 580 585
Leu Gln Glu Ser Asp Arg His Pro Asn Val Leu Arg Tyr Phe Cys
590 595 600
Thr Glu Arg Gly Pro Gln Phe His Tyr Ile Ala Leu Glu Leu Cys
605 610 615
Arg Ala Ser Leu Gln Glu Tyr Val Glu Asn Pro Asp Leu Asp Arg
620 625 630
Gly Gly Leu Glu Pro Glu Val Val Leu Gln Gln Leu Met Ser Gly
635 640 645
Leu Ala His Leu His Ser Leu His Ile Val His Arg Asp Leu Lys
650 655 660
Pro Gly Asn Ile Leu Ile Thr Gly Pro Asp Ser Gln Gly Leu Gly
665 670 675
Arg Val Val Leu Ser Asp Phe Gly Leu Cys Lys Lys Leu Pro Ala
680 685 690
Gly Arg Cys Ser Phe Ser Leu His Ser Gly Ile Pro Gly Thr Glu
695 700 705
Gly Trp Met Ala Pro Glu Leu Leu Gln Leu Leu Pro Pro Asp Ser
710 715 720
Pro Thr Ser Ala Val Asp Ile Phe Ser Ala Gly Cys Val Phe Tyr
725 730 735
Tyr Val Leu Ser Gly Gly Ser His Pro Phe Gly Asp Ser Leu Tyr
740 745 750
Arg Gln Ala Asn Ile Leu Thr Gly Ala Pro Cys Leu Ala His Leu
755 760 765
Glu Glu Glu Val His Asp Lys Val Val Ala Arg Asp Leu Val Gly
770 775 780
Ala Met Leu Ser Pro Leu Pro Gln Pro Arg Pro Ser Ala Pro Gln
785 790 795
Val Leu Ala His Pro Phe Phe Trp Ser Arg Ala Lys Gln Leu Gln
800 805 810
Phe Phe Gln Asp Val Ser Asp Trp Leu Glu Lys Glu Ser Glu Gln
815 820 825
Glu Pro Leu Val Arg Ala Leu Glu Ala Gly Gly Cys Ala Val Val
830 835 840
Arg Asp Asn Trp His Glu His Ile Ser Met Pro Leu Gln Thr Asp
845 850 855
Leu Arg Lys Phe Arg Ser Tyr Lys Gly Thr Ser Val Arg Asp Leu
860 865 870
Leu Arg Ala Val Arg Asn Lys Lys His His Tyr Arg Glu Leu Pro
875 880 885
Val Glu Val Arg Gln Ala Leu Gly Gln Val Pro Asp Gly Phe Val
890 895 900
Gln Tyr Phe Thr Asn Arg Phe Pro Arg Leu Leu Leu His Thr His
905 910 915
Arg Ala Met Arg Ser Cys Ala Ser Glu Ser Leu Phe Leu Pro Tyr
920 925 930
Tyr Pro Pro Asp Ser Glu Ala Arg Arg Pro Cys Pro Gly Ala Thr

935 940 945
Gly Arg

<210> 15
<211> 641
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7163416CD1

<400> 15
Met Phe Arg Lys Lys Lys Lys Lys Arg Pro Glu Ile Ser Ala Pro
1 5 10 15
Gln Asn Phe Gln His Arg Val His Thr Ser Phe Asp Pro Lys Glu
20 25 30
Gly Lys Phe Val Gly Leu Pro Pro Gln Trp Gln Asn Ile Leu Asp
35 40 45
Thr Leu Arg Arg Pro Lys Pro Val Val Asp Pro Ser Arg Ile Thr
50 55 60
Arg Val Gln Leu Gln Pro Met Lys Thr Val Val Arg Gly Ser Ala
65 70 75
Met Pro Val Asp Gly Tyr Ile Ser Gly Leu Leu Asn Asp Ile Gln
80 85 90
Lys Leu Ser Val Ile Ser Ser Asn Thr Leu Arg Gly Arg Ser Pro
95 100 105
Thr Ser Arg Arg Arg Ala Gln Ser Leu Gly Leu Leu Gly Asp Glu
110 115 120
His Trp Ala Thr Asp Pro Asp Met Tyr Leu Gln Ser Pro Gln Ser
125 130 135
Glu Arg Thr Asp Pro His Gly Leu Tyr Leu Ser Cys Asn Gly Gly
140 145 150
Thr Pro Ala Gly His Lys Gln Met Pro Trp Pro Glu Pro Gln Ser
155 160 165
Pro Arg Val Leu Pro Asn Gly Leu Ala Ala Lys Ala Gln Ser Leu
170 175 180
Gly Pro Ala Glu Phe Gln Gly Ala Ser Gln Arg Cys Leu Gln Leu
185 190 195
Gly Ala Cys Leu Gln Ser Ser Pro Pro Gly Ala Ser Pro Pro Thr
200 205 210
Gly Thr Asn Arg His Gly Met Lys Ala Ala Lys His Gly Ser Glu
215 220 225
Glu Ala Arg Pro Gln Ser Cys Leu Val Gly Ser Ala Thr Gly Arg
230 235 240
Pro Gly Gly Glu Gly Ser Pro Ser Pro Lys Thr Arg Glu Ser Ser
245 250 255
Leu Lys Arg Arg Leu Phe Arg Ser Met Phe Leu Ser Thr Ala Ala
260 265 270
Thr Ala Pro Pro Ser Ser Ser Lys Pro Gly Pro Pro Pro Gln Ser
275 280 285
Lys Pro Asn Ser Ser Phe Arg Pro Pro Gln Lys Asp Asn Pro Pro
290 295 300
Ser Leu Val Ala Lys Ala Gln Ser Leu Pro Ser Asp Gln Pro Val
305 310 315
Gly Thr Phe Ser Pro Leu Thr Thr Ser Asp Thr Ser Ser Pro Gln
320 325 330
Lys Ser Leu Arg Thr Ala Pro Ala Thr Gly Gln Leu Pro Gly Arg
335 340 345
Ser Ser Pro Ala Gly Ser Pro Arg Thr Trp His Ala Gln Ile Ser
350 355 360
Thr Ser Asn Leu Tyr Leu Pro Gln Asp Pro Thr Val Ala Lys Gly
365 370 375
Ala Leu Ala Gly Glu Asp Thr Gly Val Val Thr His Glu Gln Phe
380 385 390
Lys Ala Ala Leu Arg Met Val Val Asp Gln Gly Asp Pro Arg Leu

```
                    395                     400                     405
Leu Leu Asp Ser Tyr Val Lys Ile Gly Glu Gly Ser Thr Gly Ile
                410                     415                     420
Val Cys Leu Ala Arg Glu Lys His Ser Gly Arg Gln Val Ala Val
                425                     430                     435
Lys Met Met Asp Leu Arg Lys Gln Gln Arg Arg Glu Leu Leu Phe
                440                     445                     450
Asn Glu Val Val Ile Met Arg Asp Tyr Gln His Phe Asn Val Val
                455                     460                     465
Glu Met Tyr Lys Ser Tyr Leu Val Gly Glu Glu Leu Trp Val Leu
                470                     475                     480
Met Glu Phe Leu Gln Gly Gly Ala Leu Thr Asp Ile Val Ser Gln
                485                     490                     495
Val Arg Leu Asn Glu Glu Gln Ile Ala Thr Val Cys Glu Ala Val
                500                     505                     510
Leu Gln Ala Leu Ala Tyr Leu His Ala Gln Gly Val Ile His Arg
                515                     520                     525
Asp Ile Lys Ser Asp Ser Ile Leu Leu Thr Leu Asp Gly Arg Val
                530                     535                     540
Lys Leu Ser Asp Phe Gly Phe Cys Ala Gln Ile Ser Lys Asp Val
                545                     550                     555
Pro Lys Arg Lys Ser Leu Val Gly Thr Pro Tyr Trp Met Ala Pro
                560                     565                     570
Glu Val Ile Ser Arg Ser Leu Tyr Ala Thr Glu Val Asp Ile Trp
                575                     580                     585
Ser Leu Gly Ile Met Val Ile Glu Met Val Asp Gly Glu Pro Pro
                590                     595                     600
Tyr Phe Ser Asp Ser Pro Val Gln Ala Met Lys Arg Leu Arg Asp
                605                     610                     615
Ser Pro Pro Pro Lys Leu Lys Asn Ser His Lys Val Ser Trp His
                620                     625                     630
Thr Arg Val Arg Pro Arg Arg Pro His Ser Ser
                635                     640
```

<210> 16
<211> 576
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472822CD1

<400> 16

```
Met Pro Ala Leu Ser Thr Gly Ser Gly Ser Asp Thr Gly Leu Tyr
 1              5                   10                  15
Glu Leu Leu Ala Ala Leu Pro Ala Gln Leu Gln Pro His Val Asp
                20                  25                  30
Ser Gln Glu Asp Leu Thr Phe Leu Trp Asp Met Phe Gly Glu Lys
                35                  40                  45
Ser Leu His Ser Leu Val Lys Ile His Glu Lys Leu His Tyr Tyr
                50                  55                  60
Glu Lys Gln Ser Pro Val Pro Ile Leu His Gly Ala Ala Ala Leu
                65                  70                  75
Ala Asp Asp Leu Ala Glu Glu Leu Gln Asn Lys Pro Leu Asn Ser
                80                  85                  90
Glu Ile Arg Glu Leu Leu Lys Leu Leu Ser Lys Pro Asn Val Lys
                95                  100                 105
Ala Leu Leu Ser Val His Asp Thr Val Ala Gln Lys Asn Tyr Asp
                110                 115                 120
Pro Val Leu Pro Pro Met Pro Glu Asp Ile Asp Asp Glu Glu Asp
                125                 130                 135
Ser Val Lys Ile Ile Arg Leu Val Lys Asn Arg Glu Pro Leu Gly
                140                 145                 150
Ala Thr Ile Lys Lys Asp Glu Gln Thr Gly Ala Ile Ile Val Ala
                155                 160                 165
Arg Ile Met Arg Gly Gly Ala Ala Asp Arg Ser Gly Leu Ile His
```

```
                    170                          175                          180
Val Gly Asp Glu Leu Arg Glu Val Asn Gly Ile Pro Val Glu Asp
                    185                          190                          195
Lys Arg Pro Glu Glu Ile Ile Gln Ile Leu Ala Gln Ser Gln Gly
                    200                          205                          210
Ala Ile Thr Phe Lys Ile Ile Pro Gly Ser Lys Glu Glu Thr Pro
                    215                          220                          225
Ser Lys Glu Gly Lys Met Phe Ile Lys Ala Leu Phe Asp Tyr Asn
                    230                          235                          240
Pro Asn Glu Asp Lys Ala Ile Pro Cys Lys Glu Ala Gly Leu Ser
                    245                          250                          255
Phe Lys Lys Gly Asp Ile Leu Gln Ile Met Ser Gln Asp Asp Ala
                    260                          265                          270
Thr Trp Trp Gln Ala Lys His Glu Ala Asp Ala Asn Pro Arg Ala
                    275                          280                          285
Gly Leu Ile Pro Ser Lys His Phe Gln Glu Arg Arg Leu Ala Leu
                    290                          295                          300
Arg Arg Pro Glu Ile Leu Val Gln Pro Leu Lys Val Ser Asn Arg
                    305                          310                          315
Lys Ser Ser Gly Phe Arg Lys Ser Phe Arg Leu Ser Arg Lys Asp
                    320                          325                          330
Lys Lys Thr Asn Lys Ser Met Tyr Glu Cys Lys Lys Ser Asp Gln
                    335                          340                          345
Tyr Asp Thr Ala Asp Val Pro Thr Tyr Glu Glu Val Thr Pro Tyr
                    350                          355                          360
Arg Arg Gln Thr Asn Glu Lys Tyr Arg Leu Val Val Leu Val Gly
                    365                          370                          375
Pro Val Gly Val Gly Leu Asn Glu Leu Lys Arg Lys Leu Leu Ile
                    380                          385                          390
Ser Asp Thr Gln His Tyr Gly Val Thr Val Pro His Thr Thr Arg
                    395                          400                          405
Ala Arg Arg Ser Gln Glu Ser Asp Gly Val Glu Tyr Ile Phe Ile
                    410                          415                          420
Ser Lys His Leu Phe Glu Thr Asp Val Gln Asn Asn Lys Phe Ile
                    425                          430                          435
Glu Tyr Gly Glu Tyr Lys Asn Asn Tyr Tyr Gly Thr Ser Ile Asp
                    440                          445                          450
Ser Val Arg Ser Val Leu Ala Lys Asn Lys Val Cys Leu Leu Asp
                    455                          460                          465
Val Gln Pro His Thr Val Lys His Leu Arg Thr Leu Glu Phe Lys
                    470                          475                          480
Pro Tyr Val Ile Phe Ile Lys Pro Pro Ser Ile Glu Arg Leu Arg
                    485                          490                          495
Glu Thr Arg Lys Asn Ala Lys Ile Ile Ser Ser Arg Asp Asp Gln
                    500                          505                          510
Gly Ala Ala Lys Pro Phe Thr Glu Glu Asp Phe Gln Glu Met Ile
                    515                          520                          525
Lys Ser Ala Gln Ile Met Glu Ser Gln Tyr Gly His Leu Phe Asp
                    530                          535                          540
Lys Ile Ile Ile Asn Asp Asp Leu Thr Val Ala Phe Asn Glu Leu
                    545                          550                          555
Lys Thr Thr Phe Asp Lys Leu Glu Thr Glu Thr His Trp Val Pro
                    560                          565                          570
Val Ser Trp Leu His Ser
                    575
```

<210> 17
<211> 794
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477486CD1

<400> 17
Met Val Ala Gly Leu Thr Leu Gly Lys Gly Pro Glu Ser Pro Asp

```
1               5                   10                  15
Gly Asp Val Ser Val Pro Glu Arg Lys Asp Glu Val Ala Gly Gly
                20                  25                  30
Gly Gly Glu Glu Glu Glu Ala Glu Glu Arg Gly Arg His Ala Gln
                35                  40                  45
Tyr Val Gly Pro Tyr Arg Leu Glu Lys Thr Leu Gly Lys Gly Gln
                50                  55                  60
Thr Gly Leu Val Lys Leu Gly Val His Cys Ile Thr Gly Gln Lys
                65                  70                  75
Val Ala Ile Lys Ile Val Asn Arg Glu Lys Leu Ser Glu Ser Val
                80                  85                  90
Leu Met Lys Val Glu Arg Glu Ile Ala Ile Leu Lys Leu Ile Glu
                95                  100                 105
His Pro His Val Leu Lys Leu His Asp Val Tyr Glu Asn Lys Lys
                110                 115                 120
Tyr Leu Tyr Leu Val Leu Glu His Val Ser Gly Gly Glu Leu Phe
                125                 130                 135
Asp Tyr Leu Val Lys Lys Gly Arg Leu Thr Pro Lys Glu Ala Arg
                140                 145                 150
Lys Phe Phe Arg Gln Ile Val Ser Ala Leu Asp Phe Cys His Ser
                155                 160                 165
Tyr Ser Ile Cys His Arg Asp Leu Lys Pro Glu Asn Leu Leu Leu
                170                 175                 180
Asp Glu Lys Asn Asn Ile Arg Ile Ala Asp Phe Gly Met Ala Ser
                185                 190                 195
Leu Gln Val Gly Asp Ser Leu Leu Glu Thr Ser Cys Gly Ser Pro
                200                 205                 210
His Tyr Ala Cys Pro Glu Val Ile Lys Gly Glu Lys Tyr Asp Gly
                215                 220                 225
Arg Arg Ala Asp Met Trp Ser Cys Gly Val Ile Leu Phe Ala Leu
                230                 235                 240
Leu Val Gly Ala Leu Pro Phe Asp Asp Asp Asn Leu Arg Gln Leu
                245                 250                 255
Leu Glu Lys Val Lys Arg Gly Val Phe His Met Pro His Phe Ile
                260                 265                 270
Pro Pro Asp Cys Gln Ser Leu Leu Arg Gly Met Ile Glu Val Glu
                275                 280                 285
Pro Glu Lys Arg Leu Ser Leu Glu Gln Ile Gln Lys His Pro Trp
                290                 295                 300
Tyr Leu Gly Gly Lys His Glu Pro Asp Pro Cys Leu Glu Pro Ala
                305                 310                 315
Pro Gly Arg Arg Val Ala Met Arg Ser Leu Pro Ser Asn Gly Glu
                320                 325                 330
Leu Asp Pro Asp Val Leu Glu Ser Met Ala Ser Leu Gly Cys Phe
                335                 340                 345
Arg Asp Arg Glu Arg Leu His Arg Glu Leu Arg Ser Glu Glu Glu
                350                 355                 360
Asn Gln Glu Lys Met Ile Tyr Tyr Leu Leu Leu Asp Arg Lys Glu
                365                 370                 375
Arg Tyr Pro Ser Cys Glu Asp Gln Asp Leu Pro Pro Arg Asn Asp
                380                 385                 390
Val Asp Pro Pro Arg Lys Arg Val Asp Ser Pro Met Leu Ser Arg
                395                 400                 405
His Gly Lys Arg Arg Pro Glu Arg Lys Ser Met Glu Val Leu Ser
                410                 415                 420
Ile Thr Asp Ala Gly Gly Gly Gly Ser Pro Val Pro Thr Arg Arg
                425                 430                 435
Ala Leu Glu Met Ala Gln His Ser Gln Arg Ser Arg Ser Val Ser
                440                 445                 450
Gly Ala Ser Thr Gly Leu Ser Ser Ser Pro Leu Ser Ser Pro Arg
                455                 460                 465
Ser Pro Val Phe Ser Phe Ser Pro Glu Pro Gly Ala Gly Asp Glu
                470                 475                 480
Ala Arg Gly Gly Gly Ser Pro Thr Ser Lys Thr Gln Thr Leu Pro
                485                 490                 495
Ser Arg Gly Pro Arg Gly Gly Gly Ala Gly Glu Gln Pro Pro Pro
                500                 505                 510
```

```
Pro Ser Ala Arg Ser Thr Pro Leu Pro Gly Pro Pro Gly Ser Pro
                515                 520                 525
Arg Ser Ser Gly Gly Thr Pro Leu His Ser Pro Leu His Thr Pro
                530                 535                 540
Arg Ala Ser Pro Thr Gly Thr Pro Gly Thr Thr Pro Pro Pro Ser
                545                 550                 555
Pro Gly Gly Gly Val Gly Gly Ala Ala Trp Arg Ser Arg Leu Asn
                560                 565                 570
Ser Ile Arg Asn Ser Phe Leu Gly Ser Pro Arg Phe His Arg Arg
                575                 580                 585
Lys Met Gln Val Pro Thr Ala Glu Glu Met Ser Ser Leu Thr Pro
                590                 595                 600
Glu Ser Ser Pro Glu Leu Ala Lys Arg Ser Trp Phe Gly Asn Phe
                605                 610                 615
Ile Ser Leu Asp Lys Glu Glu Gln Ile Phe Leu Val Leu Lys Asp
                620                 625                 630
Lys Pro Leu Ser Ser Ile Lys Ala Asp Ile Val His Ala Phe Leu
                635                 640                 645
Ser Ile Pro Ser Leu Ser His Ser Val Leu Ser Gln Thr Ser Phe
                650                 655                 660
Arg Ala Glu Tyr Lys Ala Ser Gly Gly Pro Ser Val Phe Gln Lys
                665                 670                 675
Pro Val Arg Phe Gln Val Asp Ile Ser Ser Ser Glu Gly Pro Glu
                680                 685                 690
Pro Ser Pro Arg Arg Asp Gly Ser Gly Gly Gly Gly Ile Tyr Ser
                695                 700                 705
Val Thr Phe Thr Leu Ile Ser Gly Pro Ser Arg Arg Phe Lys Arg
                710                 715                 720
Val Val Glu Thr Ile Gln Ala Gln Leu Leu Ser Thr His Asp Gln
                725                 730                 735
Pro Ser Val Gln Ala Leu Ala Asp Glu Lys Asn Gly Ala Gln Thr
                740                 745                 750
Arg Pro Ala Gly Ala Pro Pro Arg Ser Leu Gln Pro Pro Pro Gly
                755                 760                 765
Arg Pro Asp Pro Glu Leu Ser Ser Ser Pro Arg Arg Gly Pro Pro
                770                 775                 780
Lys Asp Lys Lys Leu Leu Ala Thr Asn Gly Thr Pro Leu Pro
                785                 790

<210> 18
<211> 504
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3773709CD1

<400> 18
Met Ser Gly Leu Leu Thr Asp Pro Glu Gln Arg Ala Gln Glu Pro
  1               5                  10                  15
Arg Tyr Pro Gly Phe Val Leu Gly Leu Asp Val Gly Ser Ser Val
                 20                  25                  30
Ile Arg Cys His Val Tyr Asp Arg Ala Ala Arg Val Cys Gly Ser
                 35                  40                  45
Ser Val Gln Lys Val Glu Asn Leu Tyr Pro Gln Ile Gly Trp Val
                 50                  55                  60
Glu Ile Asp Pro Asp Val Leu Trp Ile Gln Phe Val Ala Val Ile
                 65                  70                  75
Lys Glu Ala Val Lys Ala Ala Gly Ile Gln Met Asn Gln Ile Val
                 80                  85                  90
Gly Leu Gly Ile Ser Thr Gln Arg Ala Thr Phe Ile Thr Trp Asn
                 95                 100                 105
Lys Lys Thr Gly Asn His Phe His Asn Phe Ile Ser Trp Gln Asp
                110                 115                 120
Leu Arg Ala Val Glu Leu Val Lys Ser Trp Asn Asn Ser Leu Leu
                125                 130                 135
```

```
Met Lys Ile Phe His Ser Ser Cys Arg Val Leu His Phe Phe Thr
                140                 145                 150
Arg Ser Lys Arg Leu Phe Thr Ala Ser Leu Phe Thr Phe Thr Thr
                155                 160                 165
Gln Gln Thr Ser Leu Arg Leu Val Trp Ile Leu Gln Asn Leu Thr
                170                 175                 180
Glu Val Gln Lys Ala Val Glu Glu Glu Asn Cys Cys Phe Gly Thr
                185                 190                 195
Ile Asp Thr Trp Trp Leu Tyr Lys Leu Thr Lys Gly Ser Val Tyr
                200                 205                 210
Ala Thr Asp Phe Ser Asn Ala Ser Thr Thr Gly Leu Phe Asp Pro
                215                 220                 225
Tyr Ser His Asn Phe Gly Ser Val Asp Glu Glu Ile Phe Gly Val
                230                 235                 240
Pro Ile Pro Ile Val Ala Leu Val Ala Asp Gln Gln Ser Ala Met
                245                 250                 255
Phe Gly Glu Cys Cys Phe Gln Thr Gly Asp Val Lys Leu Thr Met
                260                 265                 270
Gly Thr Gly Thr Phe Leu Asp Ile Asn Thr Gly Asn Ser Leu Gln
                275                 280                 285
Gln Thr Thr Gly Gly Phe Tyr Pro Leu Ile Gly Trp Lys Ile Gly
                290                 295                 300
Gln Glu Val Val Cys Leu Ala Glu Ser Asn Ala Gly Asp Thr Gly
                305                 310                 315
Thr Ala Ile Lys Trp Ala Gln Gln Leu Asp Leu Phe Thr Asp Ala
                320                 325                 330
Ala Glu Thr Glu Lys Met Ala Lys Ser Leu Glu Asp Ser Glu Gly
                335                 340                 345
Val Cys Phe Val Pro Ser Phe Ser Gly Leu Gln Ala Pro Leu Asn
                350                 355                 360
Asp Pro Trp Ala Cys Ala Ser Phe Met Gly Leu Lys Pro Ser Thr
                365                 370                 375
Ser Lys Tyr His Leu Val Arg Ala Ile Leu Glu Ser Ile Ala Phe
                380                 385                 390
Arg Asn Lys Gln Leu Tyr Glu Met Met Lys Lys Glu Ile His Ile
                395                 400                 405
Pro Val Arg Lys Ile Arg Ala Asp Gly Gly Val Cys Lys Asn Gly
                410                 415                 420
Phe Val Met Gln Met Thr Ser Asp Leu Ile Asn Glu Asn Ile Asp
                425                 430                 435
Arg Pro Ala Asp Ile Asp Met Ser Cys Leu Gly Ala Ala Ser Leu
                440                 445                 450
Ala Gly Leu Ala Val Gly Phe Trp Thr Asp Lys Glu Glu Leu Lys
                455                 460                 465
Lys Leu Arg Gln Ser Glu Val Val Phe Lys Pro Gln Lys Lys Cys
                470                 475                 480
Gln Glu Tyr Glu Met Ser Leu Glu Asn Trp Ala Lys Ala Val Lys
                485                 490                 495
Arg Ser Met Asn Trp Tyr Asn Lys Thr
                500

<210> 19
<211> 553
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477204CD1

<400> 19
Met Val Asp Met Gly Ala Leu Asp Asn Leu Ile Ala Asn Thr Ala
 1               5                  10                  15
Tyr Leu Gln Ala Arg Lys Pro Ser Asp Cys Asp Ser Lys Glu Leu
                20                  25                  30
Gln Arg Arg Arg Arg Ser Leu Ala Leu Pro Gly Leu Gln Gly Cys
                35                  40                  45
```

Ala Glu Leu Arg Gln Lys Leu Ser Leu Asn Phe His Ser Leu Cys
50 55 60
Glu Gln Gln Pro Ile Gly Arg Arg Leu Phe Arg Asp Phe Leu Ala
65 70 75
Thr Val Pro Thr Phe Arg Lys Ala Ala Thr Phe Leu Glu Asp Val
80 85 90
Gln Asn Trp Glu Leu Ala Glu Glu Gly Pro Thr Lys Asp Ser Ala
95 100 105
Leu Gln Gly Leu Val Ala Thr Cys Ala Ser Ala Pro Ala Pro Gly
110 115 120
Asn Pro Gln Pro Phe Leu Ser Gln Ala Val Ala Thr Lys Cys Gln
125 130 135
Ala Ala Thr Thr Glu Glu Glu Arg Val Ala Ala Val Thr Leu Ala
140 145 150
Lys Ala Glu Ala Met Ala Phe Leu Gln Glu Gln Pro Phe Lys Asp
155 160 165
Phe Val Thr Ser Ala Phe Tyr Asp Lys Phe Leu Gln Trp Lys Leu
170 175 180
Phe Glu Met Gln Pro Val Ser Asp Lys Tyr Phe Thr Glu Phe Arg
185 190 195
Val Leu Gly Lys Gly Gly Phe Gly Glu Val Cys Ala Val Gln Val
200 205 210
Lys Asn Thr Gly Lys Met Tyr Ala Cys Lys Lys Leu Asp Lys Lys
215 220 225
Arg Leu Lys Lys Lys Gly Gly Glu Lys Met Ala Leu Leu Glu Lys
230 235 240
Glu Ile Leu Glu Lys Val Ser Ser Pro Phe Ile Val Ser Leu Ala
245 250 255
Tyr Ala Phe Glu Ser Lys Thr His Leu Cys Leu Val Met Ser Leu
260 265 270
Met Asn Gly Gly Asp Leu Lys Phe His Ile Tyr Asn Val Gly Thr
275 280 285
Arg Gly Leu Asp Met Ser Arg Val Ile Phe Tyr Ser Ala Gln Ile
290 295 300
Ala Cys Gly Met Leu His Leu His Glu Leu Gly Ile Val Tyr Arg
305 310 315
Asp Met Lys Pro Glu Asn Val Leu Leu Asp Asp Leu Gly Asn Cys
320 325 330
Arg Leu Ser Asp Leu Gly Leu Ala Val Glu Met Lys Gly Gly Lys
335 340 345
Pro Ile Thr Gln Arg Ala Gly Thr Asn Gly Tyr Met Ala Pro Glu
350 355 360
Ile Leu Met Glu Lys Val Ser Tyr Ser Tyr Pro Val Asp Trp Phe
365 370 375
Ala Met Gly Cys Ser Ile Tyr Glu Met Val Ala Gly Arg Thr Pro
380 385 390
Phe Lys Asp Tyr Lys Glu Lys Val Ser Lys Glu Asp Leu Lys Gln
395 400 405
Arg Thr Leu Gln Asp Glu Val Lys Phe Gln His Asp Asn Phe Thr
410 415 420
Glu Glu Ala Lys Asp Ile Cys Arg Leu Phe Leu Ala Lys Lys Pro
425 430 435
Glu Gln Arg Leu Gly Ser Arg Glu Lys Ser Asp Asp Pro Arg Lys
440 445 450
His His Phe Phe Lys Thr Ile Asn Phe Pro Arg Leu Glu Ala Gly
455 460 465
Leu Ile Glu Pro Pro Phe Val Pro Asp Pro Ser Val Val Tyr Ala
470 475 480
Lys Asp Ile Ala Glu Ile Asp Asp Phe Ser Glu Val Arg Gly Val
485 490 495
Glu Phe Asp Asp Lys Asp Lys Gln Phe Phe Lys Asn Phe Ala Thr
500 505 510
Gly Ala Val Pro Ile Ala Trp Gln Glu Glu Ile Ile Glu Thr Gly
515 520 525
Leu Phe Glu Glu Leu Asn Asp Pro Asn Arg Pro Thr Gly Cys Glu
530 535 540
Glu Gly Asn Ser Ser Lys Ser Gly Val Cys Leu Leu Leu

545 550

<210> 20
<211> 871
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3016969CD1

<400> 20
Met Gly Pro Gly Asp Ile Ser Leu Pro Gly Arg Pro Lys Pro Gly
1 5 10 15
Pro Cys Ser Ser Pro Gly Ser Ala Ser Gln Ala Ser Ser Ser Gln
20 25 30
Val Ser Ser Leu Arg Val Gly Ser Ser Gln Val Gly Thr Glu Pro
35 40 45
Gly Pro Ser Leu Asp Ala Glu Gly Trp Thr Gln Glu Ala Glu Asp
50 55 60
Leu Ser Asp Ser Thr Pro Thr Leu Gln Arg Pro Gln Glu Gln Val
65 70 75
Thr Met Arg Lys Phe Ser Leu Gly Gly Arg Gly Gly Tyr Ala Gly
80 85 90
Val Ala Gly Tyr Gly Thr Phe Ala Phe Gly Gly Asp Ala Gly Gly
95 100 105
Met Leu Gly Gln Gly Pro Met Trp Ala Arg Ile Ala Trp Ala Val
110 115 120
Ser Gln Ser Glu Glu Glu Glu Gln Glu Glu Ala Arg Ala Glu Ser
125 130 135
Gln Ser Glu Glu Gln Gln Glu Ala Arg Ala Glu Ser Pro Leu Pro
140 145 150
Gln Val Ser Ala Arg Pro Val Pro Glu Val Gly Arg Ala Pro Thr
155 160 165
Arg Ser Ser Pro Glu Pro Thr Pro Trp Glu Asp Ile Gly Gln Val
170 175 180
Ser Leu Val Gln Ile Arg Asp Leu Ser Gly Asp Ala Glu Ala Ala
185 190 195
Asp Thr Ile Ser Leu Asp Ile Ser Glu Val Asp Pro Ala Tyr Leu
200 205 210
Asn Leu Ser Asp Leu Tyr Asp Ile Lys Tyr Leu Pro Phe Glu Phe
215 220 225
Met Ile Phe Arg Lys Val Pro Lys Ser Ala Gln Pro Glu Pro Pro
230 235 240
Ser Pro Met Ala Glu Glu Glu Leu Ala Glu Phe Pro Glu Pro Thr
245 250 255
Trp Pro Trp Pro Gly Glu Leu Gly Pro His Ala Gly Leu Glu Ile
260 265 270
Thr Glu Glu Ser Glu Asp Val Asp Ala Leu Leu Ala Glu Ala Ala
275 280 285
Val Gly Arg Lys Arg Lys Trp Ser Ser Pro Ser Arg Ser Leu Phe
290 295 300
His Phe Pro Gly Arg His Leu Pro Leu Asp Glu Pro Ala Glu Leu
305 310 315
Gly Leu Arg Glu Arg Val Lys Ala Ser Val Glu His Ile Ser Arg
320 325 330
Ile Leu Lys Gly Arg Pro Glu Gly Leu Glu Lys Glu Gly Pro Pro
335 340 345
Arg Lys Lys Pro Gly Leu Ala Ser Phe Arg Leu Ser Gly Leu Lys
350 355 360
Ser Trp Asp Arg Ala Pro Thr Phe Leu Arg Glu Leu Ser Asp Glu
365 370 375
Thr Val Val Leu Gly Gln Ser Val Thr Leu Ala Cys Gln Val Ser
380 385 390
Ala Gln Pro Ala Ala Gln Ala Thr Trp Ser Lys Asp Gly Ala Pro
395 400 405
Leu Glu Ser Ser Ser Arg Val Leu Ile Ser Ala Thr Leu Lys Asn

410 415 420
Phe Gln Leu Leu Thr Ile Leu Val Val Val Ala Glu Asp Leu Gly
425 430 435
Val Tyr Thr Cys Ser Val Ser Asn Ala Leu Gly Thr Val Thr Thr
440 445 450
Thr Gly Val Leu Arg Lys Ala Glu Arg Pro Ser Ser Ser Pro Cys
455 460 465
Pro Asp Ile Gly Glu Val Tyr Ala Asp Gly Val Leu Leu Val Trp
470 475 480
Lys Pro Val Glu Ser Tyr Gly Pro Val Thr Tyr Ile Val Gln Cys
485 490 495
Ser Leu Glu Gly Gly Ser Trp Thr Thr Leu Ala Ser Asp Ile Phe
500 505 510
Asp Cys Cys Tyr Leu Thr Ser Lys Leu Ser Arg Gly Gly Thr Tyr
515 520 525
Thr Phe Arg Thr Ala Cys Val Ser Lys Ala Gly Met Gly Pro Tyr
530 535 540
Ser Ser Pro Ser Glu Gln Val Leu Leu Gly Gly Pro Ser His Leu
545 550 555
Ala Ser Glu Glu Glu Ser Gln Gly Arg Ser Ala Gln Pro Leu Pro
560 565 570
Ser Thr Lys Thr Phe Ala Phe Gln Thr Gln Ile Gln Arg Gly Arg
575 580 585
Phe Ser Val Val Arg Gln Cys Trp Glu Lys Ala Ser Gly Arg Ala
590 595 600
Leu Ala Ala Lys Ile Ile Pro Tyr His Pro Lys Asp Lys Thr Ala
605 610 615
Val Leu Arg Glu Tyr Glu Ala Leu Lys Gly Leu Arg His Pro His
620 625 630
Leu Ala Gln Leu His Ala Ala Tyr Leu Ser Pro Arg His Leu Val
635 640 645
Leu Ile Leu Glu Leu Cys Ser Gly Pro Glu Leu Leu Pro Cys Leu
650 655 660
Ala Glu Arg Ala Ser Tyr Ser Glu Ser Glu Val Lys Asp Tyr Leu
665 670 675
Trp Gln Met Leu Ser Ala Thr Gln Tyr Leu His Asn Gln His Ile
680 685 690
Leu His Leu Asp Leu Arg Ser Glu Asn Met Ile Ile Thr Glu Tyr
695 700 705
Asn Leu Leu Lys Val Val Asp Leu Gly Asn Ala Gln Ser Leu Ser
710 715 720
Gln Glu Lys Val Leu Pro Ser Asp Lys Phe Lys Asp Tyr Leu Glu
725 730 735
Thr Met Ala Pro Glu Leu Leu Glu Gly Gln Gly Ala Val Pro Gln
740 745 750
Thr Asp Ile Trp Ala Ile Gly Val Thr Ala Phe Ile Met Leu Ser
755 760 765
Ala Glu Tyr Pro Val Ser Ser Glu Gly Ala Arg Asp Leu Gln Arg
770 775 780
Gly Leu Arg Lys Gly Leu Val Arg Leu Ser Arg Cys Tyr Ala Gly
785 790 795
Leu Ser Gly Gly Ala Val Ala Phe Leu Arg Ser Thr Leu Cys Ala
800 805 810
Gln Pro Trp Gly Arg Pro Cys Ala Ser Ser Cys Leu Gln Cys Pro
815 820 825
Trp Leu Thr Glu Glu Gly Pro Ala Cys Ser Arg Pro Ala Pro Val
830 835 840
Thr Phe Pro Thr Ala Arg Leu Arg Val Phe Val Arg Asn Arg Glu
845 850 855
Lys Arg Arg Ala Leu Leu Tyr Lys Arg His Asn Leu Ala Gln Val
860 865 870
Arg

<210> 21
<211> 765
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 063497CD1

<400> 21
```
Met Ala Gly Phe Lys Arg Gly Tyr Asp Gly Lys Ile Ala Gly Leu
  1               5                  10                  15
Tyr Asp Leu Asp Lys Thr Leu Gly Arg Gly His Phe Ala Val Val
                 20                  25                  30
Lys Leu Ala Arg His Val Phe Thr Gly Glu Lys Val Ala Val Lys
                 35                  40                  45
Val Ile Asp Lys Thr Lys Leu Asp Thr Leu Ala Thr Gly His Leu
                 50                  55                  60
Phe Gln Glu Val Arg Cys Met Lys Leu Val Gln His Pro Asn Ile
                 65                  70                  75
Val Arg Leu Tyr Glu Val Ile Asp Thr Gln Thr Lys Leu Tyr Leu
                 80                  85                  90
Ile Leu Glu Leu Gly Asp Gly Gly Asp Met Phe Asp Tyr Ile Met
                 95                 100                 105
Lys His Glu Glu Gly Leu Asn Glu Asp Leu Ala Lys Lys Tyr Phe
                110                 115                 120
Ala Gln Ile Val His Ala Ile Ser Tyr Cys His Lys Leu His Val
                125                 130                 135
Val His Arg Asp Leu Lys Pro Glu Asn Val Val Phe Phe Glu Lys
                140                 145                 150
Gln Gly Leu Val Lys Leu Thr Asp Phe Gly Phe Ser Asn Lys Phe
                155                 160                 165
Gln Pro Gly Lys Lys Leu Thr Thr Ser Cys Gly Ser Leu Ala Tyr
                170                 175                 180
Ser Ala Pro Glu Ile Leu Leu Gly Asp Glu Tyr Asp Ala Pro Ala
                185                 190                 195
Val Asp Ile Trp Ser Leu Gly Val Ile Leu Phe Met Leu Val Cys
                200                 205                 210
Gly Gln Pro Pro Phe Gln Glu Ala Asn Asp Ser Glu Thr Leu Thr
                215                 220                 225
Met Ile Met Asp Cys Lys Tyr Thr Val Pro Ser His Val Ser Lys
                230                 235                 240
Glu Cys Lys Asp Leu Ile Thr Arg Met Leu Gln Arg Asp Pro Lys
                245                 250                 255
Arg Arg Ala Ser Leu Glu Glu Ile Glu Asn His Pro Trp Leu Gln
                260                 265                 270
Gly Val Asp Pro Ser Pro Ala Thr Lys Tyr Asn Ile Pro Leu Val
                275                 280                 285
Ser Tyr Lys Asn Leu Ser Glu Glu Glu His Asn Ser Ile Ile Gln
                290                 295                 300
Arg Met Val Leu Gly Asp Ile Ala Asp Arg Asp Ala Ile Val Glu
                305                 310                 315
Ala Leu Glu Thr Asn Arg Tyr Asn His Ile Thr Ala Thr Tyr Phe
                320                 325                 330
Leu Leu Ala Glu Arg Ile Leu Arg Glu Lys Gln Glu Lys Glu Ile
                335                 340                 345
Gln Thr Arg Ser Ala Ser Pro Ser Asn Ile Lys Ala Gln Phe Arg
                350                 355                 360
Gln Ser Trp Pro Thr Lys Ile Asp Val Pro Gln Asp Leu Glu Asp
                365                 370                 375
Asp Leu Thr Ala Thr Pro Leu Ser His Ala Thr Val Pro Gln Ser
                380                 385                 390
Pro Ala Arg Ala Ala Asp Ser Val Leu Asn Gly His Arg Ser Lys
                395                 400                 405
Gly Leu Cys Asp Ser Ala Lys Lys Asp Asp Leu Pro Glu Leu Ala
                410                 415                 420
Gly Pro Ala Leu Ser Thr Val Pro Pro Ala Ser Leu Lys Pro Thr
                425                 430                 435
Ala Ser Gly Arg Lys Cys Leu Phe Arg Val Glu Glu Asp Glu Glu
                440                 445                 450
```

Glu Asp Glu Glu Asp Lys Lys Pro Met Ser Leu Ser Thr Gln Val
455 460 465
Val Leu Arg Arg Lys Pro Ser Val Thr Asn Arg Leu Thr Ser Arg
470 475 480
Lys Ser Ala Pro Val Leu Asn Gln Ile Phe Glu Glu Gly Glu Ser
485 490 495
Asp Asp Glu Phe Asp Met Asp Glu Asn Leu Pro Pro Lys Leu Ser
500 505 510
Arg Leu Lys Met Asn Ile Ala Ser Pro Gly Thr Val His Lys Arg
515 520 525
Tyr His Arg Arg Lys Ser Gln Gly Arg Gly Ser Ser Cys Ser Ser
530 535 540
Ser Glu Thr Ser Asp Asp Asp Ser Glu Ser Arg Arg Arg Leu Asp
545 550 555
Lys Asp Ser Gly Phe Thr Tyr Ser Trp His Arg Arg Asp Ser Ser
560 565 570
Glu Gly Pro Pro Gly Ser Glu Gly Asp Gly Gly Gly Gln Ser Lys
575 580 585
Pro Ser Asn Ala Ser Gly Gly Val Asp Lys Ala Ser Pro Ser Glu
590 595 600
Asn Asn Ala Gly Gly Gly Ser Pro Ser Ser Gly Ser Gly Gly Asn
605 610 615
Pro Thr Asn Thr Ser Gly Thr Thr Arg Arg Cys Ala Gly Pro Ser
620 625 630
Asn Ser Met Gln Leu Ala Ser Arg Ser Ala Gly Glu Leu Val Glu
635 640 645
Ser Leu Lys Leu Met Ser Leu Cys Leu Gly Ser Gln Leu His Gly
650 655 660
Ser Thr Lys Tyr Ile Ile Asp Pro Gln Asn Gly Leu Ser Phe Ser
665 670 675
Ser Val Lys Val Gln Glu Lys Ser Thr Trp Lys Met Cys Ile Ser
680 685 690
Ser Thr Gly Asn Ala Gly Gln Val Pro Ala Val Gly Gly Ile Lys
695 700 705
Phe Phe Ser Asp His Met Ala Asp Thr Thr Thr Glu Leu Glu Arg
710 715 720
Ile Lys Ser Lys Asn Leu Lys Asn Asn Val Leu Gln Leu Pro Leu
725 730 735
Cys Glu Lys Thr Ile Ser Val Asn Ile Gln Arg Asn Pro Lys Glu
740 745 750
Gly Leu Leu Cys Ala Ser Ser Pro Ala Ser Cys Cys His Val Ile
755 760 765

<210> 22
<211> 588
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1625436CD1

<400> 22
Met Ala Thr Thr Ala Thr Cys Thr Arg Phe Thr Asp Asp Tyr Gln
1 5 10 15
Leu Phe Glu Glu Leu Gly Lys Gly Ala Phe Ser Val Val Arg Arg
20 25 30
Cys Val Lys Lys Thr Ser Thr Gln Glu Tyr Ala Ala Lys Ile Ile
35 40 45
Asn Thr Lys Lys Leu Ser Ala Arg Asp His Gln Lys Leu Glu Arg
50 55 60
Glu Ala Arg Ile Cys Arg Leu Leu Lys His Pro Asn Ile Val Arg
65 70 75
Leu His Asp Ser Ile Ser Glu Glu Gly Phe His Tyr Leu Val Phe
80 85 90
Asp Leu Val Thr Gly Gly Glu Leu Phe Glu Asp Ile Val Ala Arg

95 100 105
Glu Tyr Tyr Ser Glu Ala Asp Ala Ser His Cys Ile His Gln Ile
110 115 120
Leu Glu Ser Val Asn His Ile His Gln His Asp Ile Val His Arg
125 130 135
Asp Leu Lys Pro Glu Asn Leu Leu Leu Ala Ser Lys Cys Lys Gly
140 145 150
Ala Ala Val Lys Leu Ala Asp Phe Gly Leu Ala Ile Glu Val Gln
155 160 165
Gly Glu Gln Gln Ala Trp Phe Gly Phe Ala Gly Thr Pro Gly Tyr
170 175 180
Leu Ser Pro Glu Val Leu Arg Lys Asp Pro Tyr Gly Lys Pro Val
185 190 195
Asp Ile Trp Ala Cys Gly Val Ile Leu Tyr Ile Leu Leu Val Gly
200 205 210
Tyr Pro Pro Phe Trp Asp Glu Asp Gln His Lys Leu Tyr Gln Gln
215 220 225
Ile Lys Ala Gly Ala Tyr Asp Phe Pro Ser Pro Glu Trp Asp Thr
230 235 240
Val Thr Pro Glu Ala Lys Asn Leu Ile Asn Gln Met Leu Thr Ile
245 250 255
Asn Pro Ala Lys Arg Ile Thr Ala Asp Gln Ala Leu Lys Tyr Pro
260 265 270
Trp Val Cys Gln Arg Ser Thr Val Ala Ser Met Met His Arg Gln
275 280 285
Glu Thr Val Glu Cys Leu Arg Lys Phe Asn Ala Arg Arg Lys Leu
290 295 300
Lys Gly Ala Ile Leu Thr Thr Met Leu Val Ser Arg Asn Phe Ser
305 310 315
Val Gly Arg Gln Ser Ser Ala Pro Ala Ser Pro Ala Ala Ser Ala
320 325 330
Ala Gly Leu Ala Gly Gln Ala Ala Lys Ser Leu Leu Asn Lys Lys
335 340 345
Ser Asp Gly Gly Val Lys Lys Arg Lys Ser Ser Ser Ser Val His
350 355 360
Leu Met Pro Gln Ser Asn Asn Lys Asn Ser Leu Val Ser Pro Ala
365 370 375
Gln Glu Pro Ala Pro Leu Gln Thr Ala Met Glu Pro Gln Thr Thr
380 385 390
Val Val His Asn Ala Thr Asp Gly Ile Lys Gly Ser Thr Glu Ser
395 400 405
Cys Asn Thr Thr Thr Glu Asp Glu Asp Leu Lys Ala Ala Pro Leu
410 415 420
Arg Thr Gly Asn Gly Ser Ser Val Pro Glu Gly Arg Ser Ser Arg
425 430 435
Asp Arg Thr Ala Pro Ser Ala Gly Met Gln Pro Gln Pro Ser Leu
440 445 450
Cys Ser Ser Ala Met Arg Lys Gln Glu Ile Ile Lys Ile Thr Glu
455 460 465
Gln Leu Ile Glu Ala Ile Asn Asn Gly Asp Phe Glu Ala Tyr Thr
470 475 480
Lys Ile Cys Asp Pro Gly Leu Thr Ser Phe Glu Pro Glu Ala Leu
485 490 495
Gly Asn Leu Val Glu Gly Met Asp Phe His Lys Phe Tyr Phe Glu
500 505 510
Asn Leu Leu Ser Lys Asn Ser Lys Pro Ile His Thr Thr Ile Leu
515 520 525
Asn Pro His Val His Val Ile Gly Glu Asp Ala Ala Cys Ile Ala
530 535 540
Tyr Ile Arg Leu Thr Gln Tyr Ile Asp Gly Gln Gly Arg Pro Arg
545 550 555
Thr Ser Gln Ser Glu Glu Thr Arg Val Trp His Arg Arg Asp Gly
560 565 570
Lys Trp Leu Asn Val His Tyr His Cys Ser Gly Ala Pro Ala Ala
575 580 585
Pro Leu Gln

<210> 23
<211> 1798
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3330646CD1

<400> 23
Met Lys Arg Ser Arg Cys Arg Asp Arg Pro Gln Pro Pro Pro Pro
1 5 10 15
Asp Arg Arg Glu Asp Gly Val Gln Arg Ala Ala Glu Leu Ser Gln
20 25 30
Ser Leu Pro Pro Arg Arg Arg Ala Pro Pro Gly Arg Gln Arg Leu
35 40 45
Glu Glu Arg Thr Gly Pro Ala Gly Pro Glu Gly Lys Glu Gln Asp
50 55 60
Val Ala Thr Gly Val Ser Pro Leu Leu Phe Arg Lys Leu Ser Asn
65 70 75
Pro Asp Ile Phe Ser Ser Thr Gly Lys Val Lys Leu Gln Arg Gln
80 85 90
Leu Ser Gln Asp Asp Cys Lys Leu Trp Arg Gly Asn Leu Ala Ser
95 100 105
Ser Leu Ser Gly Lys Gln Leu Leu Pro Leu Ser Ser Ser Val His
110 115 120
Ser Ser Val Gly Gln Val Thr Trp Gln Ser Ser Gly Glu Ala Ser
125 130 135
Asn Leu Val Arg Met Arg Asn Gln Ser Leu Gly Gln Ser Ala Pro
140 145 150
Ser Leu Thr Ala Gly Leu Lys Glu Leu Ser Leu Pro Arg Arg Gly
155 160 165
Ser Phe Cys Arg Thr Ser Asn Arg Lys Ser Leu Ile Val Thr Ser
170 175 180
Ser Thr Ser Pro Thr Leu Pro Arg Pro His Ser Pro Leu His Gly
185 190 195
His Thr Gly Asn Ser Pro Leu Asp Ser Pro Arg Asn Phe Ser Pro
200 205 210
Asn Ala Pro Ala His Phe Ser Phe Val Pro Ala Arg Arg Thr Asp
215 220 225
Gly Arg Arg Trp Ser Leu Ala Ser Leu Pro Ser Ser Gly Tyr Gly
230 235 240
Thr Asn Thr Pro Ser Ser Thr Val Ser Ser Ser Cys Ser Ser Gln
245 250 255
Glu Lys Leu His Gln Leu Pro Phe Gln Pro Thr Ala Asp Glu Leu
260 265 270
His Phe Leu Thr Lys His Phe Ser Thr Glu Ser Val Pro Asp Glu
275 280 285
Glu Gly Arg Gln Ser Pro Ala Met Arg Pro Arg Ser Arg Ser Leu
290 295 300
Ser Pro Gly Arg Ser Pro Val Ser Phe Asp Ser Glu Ile Ile Met
305 310 315
Met Asn His Val Tyr Lys Glu Arg Phe Pro Lys Ala Thr Ala Gln
320 325 330
Met Glu Glu Arg Leu Ala Glu Phe Ile Ser Ser Asn Thr Pro Asp
335 340 345
Ser Val Leu Pro Leu Ala Asp Gly Ala Leu Ser Phe Ile His His
350 355 360
Gln Val Ile Glu Met Ala Arg Asp Cys Leu Asp Lys Ser Arg Ser
365 370 375
Gly Leu Ile Thr Ser Gln Tyr Phe Tyr Glu Leu Gln Glu Asn Leu
380 385 390
Glu Lys Leu Leu Gln Asp Ala His Glu Arg Ser Glu Ser Ser Glu
395 400 405
Val Ala Phe Val Met Gln Leu Val Lys Lys Leu Met Ile Ile Ile
410 415 420
Ala Arg Pro Ala Arg Leu Leu Glu Cys Leu Glu Phe Asp Pro Glu

```
                    425                 430                 435
Glu Phe Tyr His Leu Leu Glu Ala Ala Glu Gly His Ala Lys Glu
                440                 445                 450
Gly Gln Gly Ile Lys Cys Asp Ile Pro Arg Tyr Ile Val Ser Gln
                455                 460                 465
Leu Gly Leu Thr Arg Asp Pro Leu Glu Glu Met Ala Gln Leu Ser
                470                 475                 480
Ser Cys Asp Ser Pro Asp Thr Pro Glu Thr Asp Asp Ser Ile Glu
                485                 490                 495
Gly His Gly Ala Ser Leu Pro Ser Lys Lys Thr Pro Ser Glu Glu
                500                 505                 510
Asp Phe Glu Thr Ile Lys Leu Ile Ser Asn Gly Ala Tyr Gly Ala
                515                 520                 525
Val Phe Leu Val Arg His Lys Ser Thr Arg Gln Arg Phe Ala Met
                530                 535                 540
Lys Lys Ile Asn Lys Gln Asn Leu Ile Leu Arg Asn Gln Ile Gln
                545                 550                 555
Gln Ala Phe Val Glu Arg Asp Ile Leu Thr Phe Ala Glu Asn Pro
                560                 565                 570
Phe Val Val Ser Met Phe Cys Ser Phe Asp Thr Lys Arg His Leu
                575                 580                 585
Cys Met Val Met Glu Tyr Val Glu Gly Gly Asp Cys Ala Thr Leu
                590                 595                 600
Leu Lys Asn Ile Gly Ala Leu Pro Val Asp Met Val Arg Leu Tyr
                605                 610                 615
Phe Ala Glu Thr Val Leu Ala Leu Glu Tyr Leu His Asn Tyr Gly
                620                 625                 630
Ile Val His Arg Asp Leu Lys Pro Asp Asn Leu Leu Ile Thr Ser
                635                 640                 645
Met Gly His Ile Lys Leu Thr Asp Phe Gly Leu Ser Lys Met Gly
                650                 655                 660
Leu Met Ser Leu Thr Thr Asn Leu Tyr Glu Gly His Ile Glu Lys
                665                 670                 675
Asp Ala Arg Glu Phe Leu Asp Lys Gln Val Cys Gly Thr Pro Glu
                680                 685                 690
Tyr Ile Ala Pro Glu Val Ile Leu Arg Gln Gly Tyr Gly Lys Pro
                695                 700                 705
Val Asp Trp Trp Ala Met Gly Ile Ile Leu Tyr Glu Phe Leu Val
                710                 715                 720
Gly Cys Val Pro Phe Phe Gly Asp Thr Pro Glu Glu Leu Phe Gly
                725                 730                 735
Gln Val Ile Ser Asp Glu Ile Val Trp Pro Glu Gly Asp Glu Ala
                740                 745                 750
Leu Pro Pro Asp Ala Gln Asp Leu Thr Ser Lys Leu Leu His Gln
                755                 760                 765
Asn Pro Leu Glu Arg Leu Gly Thr Gly Ser Ala Tyr Glu Val Lys
                770                 775                 780
Gln His Pro Phe Phe Thr Gly Leu Asp Trp Thr Gly Leu Leu Arg
                785                 790                 795
Gln Lys Ala Glu Phe Ile Pro Gln Leu Glu Ser Glu Asp Asp Thr
                800                 805                 810
Ser Tyr Phe Asp Thr Arg Ser Glu Arg Tyr His His Met Asp Ser
                815                 820                 825
Glu Asp Glu Glu Glu Val Ser Glu Asp Gly Cys Leu Glu Ile Arg
                830                 835                 840
Gln Phe Ser Ser Cys Ser Pro Arg Phe Asn Lys Val Tyr Ser Ser
                845                 850                 855
Met Glu Arg Leu Ser Leu Leu Glu Glu Arg Arg Thr Pro Pro Pro
                860                 865                 870
Thr Lys Arg Ser Leu Ser Glu Glu Lys Glu Asp His Ser Asp Gly
                875                 880                 885
Leu Ala Gly Leu Lys Gly Arg Asp Arg Ser Trp Val Ile Gly Ser
                890                 895                 900
Pro Glu Ile Leu Arg Lys Arg Leu Ser Val Ser Glu Ser Ser His
                905                 910                 915
Thr Glu Ser Asp Ser Ser Pro Pro Met Thr Val Arg Arg Arg Cys
                920                 925                 930
```

```
Ser Gly Leu Leu Asp Ala Pro Arg Phe Pro Glu Gly Pro Glu Glu
                935                 940                 945
Ala Ser Ser Thr Leu Arg Arg Gln Pro Gln Glu Gly Ile Trp Val
                950                 955                 960
Leu Thr Pro Pro Ser Gly Glu Gly Val Ser Gly Pro Val Thr Glu
                965                 970                 975
His Ser Gly Glu Gln Arg Pro Lys Leu Asp Glu Glu Ala Val Gly
                980                 985                 990
Arg Ser Ser Gly Ser Ser Pro Ala Met Glu Thr Arg Gly Arg Gly
                995                1000                1005
Thr Ser Gln Leu Ala Glu Gly Ala Thr Ala Lys Ala Ile Ser Asp
               1010                1015                1020
Leu Ala Val Arg Arg Ala Arg His Arg Leu Leu Ser Gly Asp Ser
               1025                1030                1035
Thr Glu Lys Arg Thr Ala Arg Pro Val Asn Lys Val Ile Lys Ser
               1040                1045                1050
Ala Ser Ala Thr Ala Leu Ser Leu Leu Ile Pro Ser Glu His His
               1055                1060                1065
Thr Cys Ser Pro Leu Ala Ser Pro Met Ser Pro His Ser Gln Ser
               1070                1075                1080
Ser Asn Pro Ser Ser Arg Asp Ser Ser Pro Ser Arg Asp Phe Leu
               1085                1090                1095
Pro Ala Leu Gly Ser Met Arg Pro Pro Ile Ile Ile His Arg Ala
               1100                1105                1110
Gly Lys Lys Tyr Gly Phe Thr Leu Arg Ala Ile Arg Val Tyr Met
               1115                1120                1125
Gly Asp Ser Asp Val Tyr Thr Val His His Met Val Trp His Val
               1130                1135                1140
Glu Asp Gly Gly Pro Ala Ser Glu Ala Gly Leu Arg Gln Gly Asp
               1145                1150                1155
Leu Ile Thr His Val Asn Gly Glu Pro Val His Gly Leu Val His
               1160                1165                1170
Thr Glu Val Val Glu Leu Ile Leu Lys Ser Gly Asn Lys Val Ala
               1175                1180                1185
Ile Ser Thr Thr Pro Leu Glu Asn Thr Ser Ile Lys Val Gly Pro
               1190                1195                1200
Ala Arg Lys Gly Ser Tyr Lys Ala Lys Met Ala Arg Arg Ser Lys
               1205                1210                1215
Arg Ser Arg Gly Lys Asp Gly Gln Glu Ser Arg Lys Arg Ser Ser
               1220                1225                1230
Leu Phe Arg Lys Ile Thr Lys Gln Ala Ser Leu Leu His Thr Ser
               1235                1240                1245
Arg Ser Leu Ser Ser Leu Asn Arg Ser Leu Ser Ser Gly Glu Ser
               1250                1255                1260
Gly Pro Gly Ser Pro Thr His Ser His Ser Leu Ser Pro Arg Ser
               1265                1270                1275
Pro Thr Gln Gly Tyr Arg Val Thr Pro Asp Ala Val His Ser Val
               1280                1285                1290
Gly Gly Asn Ser Ser Gln Ser Ser Ser Pro Ser Ser Ser Val Pro
               1295                1300                1305
Ser Ser Pro Ala Gly Ser Gly His Thr Arg Pro Ser Ser Leu His
               1310                1315                1320
Gly Leu Ala Pro Lys Leu Gln Arg Gln Tyr Arg Ser Pro Arg Arg
               1325                1330                1335
Lys Ser Ala Gly Ser Ile Pro Leu Ser Pro Leu Ala His Thr Pro
               1340                1345                1350
Ser Pro Pro Pro Pro Thr Ala Ser Pro Gln Arg Ser Pro Ser Pro
               1355                1360                1365
Leu Ser Gly His Val Ala Gln Ala Phe Pro Thr Lys Leu His Leu
               1370                1375                1380
Ser Pro Pro Leu Gly Arg Gln Leu Ser Arg Pro Lys Ser Ala Glu
               1385                1390                1395
Pro Pro Arg Ser Pro Leu Leu Lys Arg Val Gln Ser Ala Glu Lys
               1400                1405                1410
Leu Ala Ala Ala Leu Ala Ala Ser Glu Lys Lys Leu Ala Thr Ser
               1415                1420                1425
Arg Lys His Ser Leu Asp Leu Pro His Ser Glu Leu Lys Lys Glu
```

```
                1430                1435                1440
Leu Pro Pro Arg Glu Val Ser Pro Leu Glu Val Val Gly Ala Arg
                1445                1450                1455
Ser Val Leu Ser Gly Lys Gly Ala Leu Pro Gly Lys Gly Val Leu
                1460                1465                1470
Gln Pro Ala Pro Ser Arg Ala Leu Gly Thr Leu Arg Gln Asp Arg
                1475                1480                1485
Ala Glu Arg Arg Glu Ser Leu Gln Lys Gln Glu Ala Ile Arg Glu
                1490                1495                1500
Val Asp Ser Ser Glu Asp Asp Thr Glu Glu Gly Pro Glu Asn Ser
                1505                1510                1515
Gln Gly Ala Gln Glu Leu Ser Leu Ala Pro His Pro Glu Val Ser
                1520                1525                1530
Gln Ser Val Ala Pro Lys Gly Ala Gly Glu Ser Gly Glu Glu Asp
                1535                1540                1545
Pro Phe Pro Ser Arg Asp Pro Arg Ser Leu Gly Pro Met Val Pro
                1550                1555                1560
Ser Leu Leu Thr Gly Ile Thr Leu Gly Pro Pro Arg Met Glu Ser
                1565                1570                1575
Pro Ser Gly Pro His Arg Arg Leu Gly Ser Pro Gln Ala Ile Glu
                1580                1585                1590
Glu Ala Ala Ser Ser Ser Ser Ala Gly Pro Asn Leu Gly Gln Ser
                1595                1600                1605
Gly Ala Thr Asp Pro Ile Pro Pro Glu Gly Cys Trp Lys Ala Gln
                1610                1615                1620
His Leu His Thr Gln Ala Leu Thr Ala Leu Ser Pro Ser Thr Ser
                1625                1630                1635
Gly Leu Thr Pro Thr Ser Ser Cys Ser Pro Pro Ser Ser Thr Ser
                1640                1645                1650
Gly Lys Leu Ser Met Trp Ser Trp Lys Ser Leu Ile Glu Gly Pro
                1655                1660                1665
Asp Arg Ala Ser Pro Ser Arg Lys Ala Thr Met Ala Gly Gly Leu
                1670                1675                1680
Ala Asn Leu Gln Asp Leu Glu Asn Thr Thr Pro Ala Gln Pro Lys
                1685                1690                1695
Asn Leu Ser Pro Arg Glu Gln Gly Lys Thr Gln Pro Pro Ser Ala
                1700                1705                1710
Pro Arg Leu Ala His Pro Ser Tyr Glu Asp Pro Ser Gln Gly Trp
                1715                1720                1725
Leu Trp Glu Ser Glu Cys Ala Gln Ala Val Lys Glu Asp Pro Ala
                1730                1735                1740
Leu Ser Ile Thr Gln Val Pro Asp Ala Ser Gly Asp Arg Arg Gln
                1745                1750                1755
Asp Val Pro Cys Arg Gly Cys Pro Leu Thr Gln Lys Ser Glu Pro
                1760                1765                1770
Ser Leu Arg Arg Gly Gln Glu Pro Gly Gly His Gln Lys His Arg
                1775                1780                1785
Asp Leu Ala Leu Val Pro Asp Glu Leu Leu Lys Gln Thr
                1790                1795
```

```
<210> 24
<211> 362
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3562763CD1

<400> 24
Met Asp Pro Val Ala Ala Glu Ala Pro Gly Glu Ala Phe Leu Ala
 1               5                  10                  15
Arg Arg Arg Pro Glu Gly Gly Gly Gly Ser Ala Arg Pro Arg Tyr
                20                  25                  30
Ser Leu Leu Ala Glu Ile Gly Arg Gly Ser Tyr Gly Val Val Tyr
                35                  40                  45
Glu Ala Val Ala Gly Arg Ser Gly Ala Arg Val Ala Val Lys Lys
```

```
                    50                      55                      60
Ile Arg Cys Asp Ala Pro Glu Asn Val Glu Leu Ala Leu Ala Glu
                    65                      70                      75
Phe Trp Ala Leu Thr Ser Leu Lys Arg Arg His Gln Asn Val Val
                    80                      85                      90
Gln Phe Glu Glu Cys Val Leu Gln Arg Asn Gly Leu Ala Gln Arg
                    95                      100                     105
Met Ser His Gly Asn Lys Ser Ser Gln Leu Tyr Leu Arg Leu Val
                    110                     115                     120
Glu Thr Ser Leu Lys Gly Glu Arg Ile Leu Gly Tyr Ala Glu Glu
                    125                     130                     135
Pro Cys Tyr Leu Trp Phe Val Met Glu Phe Cys Glu Gly Gly Asp
                    140                     145                     150
Leu Asn Gln Tyr Val Leu Ser Arg Arg Pro Asp Pro Ala Thr Asn
                    155                     160                     165
Lys Ser Phe Met Leu Gln Leu Thr Ser Ala Ile Ala Phe Leu His
                    170                     175                     180
Lys Asn His Ile Val His Arg Asp Leu Lys Pro Asp Asn Ile Leu
                    185                     190                     195
Ile Thr Glu Arg Ser Gly Thr Pro Ile Leu Lys Val Ala Asp Phe
                    200                     205                     210
Gly Leu Ser Lys Val Cys Ala Gly Leu Ala Pro Arg Gly Lys Glu
                    215                     220                     225
Gly Asn Gln Asp Asn Lys Asn Val Asn Val Asn Lys Tyr Trp Leu
                    230                     235                     240
Ser Ser Ala Cys Gly Ser Asp Phe Tyr Met Ala Pro Glu Val Trp
                    245                     250                     255
Glu Gly His Tyr Thr Ala Lys Ala Asp Ile Phe Ala Leu Gly Ile
                    260                     265                     270
Ile Ile Trp Ala Met Ile Glu Arg Ile Thr Phe Ile Asp Ser Glu
                    275                     280                     285
Thr Lys Lys Glu Leu Leu Gly Thr Tyr Ile Lys Gln Gly Thr Glu
                    290                     295                     300
Ile Val Pro Val Gly Glu Ala Leu Leu Glu Asn Pro Lys Met Glu
                    305                     310                     315
Leu His Ile Pro Gln Lys Arg Arg Thr Ser Met Ser Glu Gly Ile
                    320                     325                     330
Lys Gln Leu Leu Lys Asp Met Leu Ala Ala Asn Pro Gln Asp Arg
                    335                     340                     345
Pro Asp Ala Phe Glu Leu Glu Thr Arg Met Asp Gln Val Thr Cys
                    350                     355                     360
Ala Ala


<210> 25
<211> 275
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 621293CD1

<400> 25
Met Val Pro Glu Asp Ile Ser Glu Leu Glu Thr Ala Gln Lys Leu
  1                 5                       10                      15
Leu Glu Tyr His Arg Asn Ile Val Arg Val Ile Pro Ser Tyr Pro
                    20                      25                      30
Lys Ile Leu Lys Val Ile Ser Ala Asp Gln Pro Cys Val Asp Val
                    35                      40                      45
Phe Tyr Gln Ala Leu Thr Tyr Val Gln Ser Asn His Arg Thr Asn
                    50                      55                      60
Ala Pro Phe Thr Pro Arg Val Leu Leu Leu Gly Pro Val Gly Ser
                    65                      70                      75
Gly Lys Ser Leu Gln Ala Ala Leu Leu Ala Gln Lys Tyr Arg Leu
                    80                      85                      90
Val Asn Val Cys Cys Gly Gln Leu Leu Lys Glu Ala Val Ala Asp
```

```
                    95                  100                 105
Arg Thr Thr Phe Gly Glu Leu Ile Gln Pro Phe Phe Glu Lys Glu
                110                 115                 120
Met Ala Val Pro Asp Ser Leu Leu Met Lys Val Leu Ser Gln Arg
                125                 130                 135
Leu Asp Gln Gln Asp Cys Ile Gln Lys Gly Trp Val Leu His Gly
                140                 145                 150
Val Pro Arg Asp Leu Asp Gln Ala His Leu Leu Asn Arg Leu Gly
                155                 160                 165
Tyr Asn Pro Asn Arg Val Phe Phe Leu Asn Val Pro Phe Asp Ser
                170                 175                 180
Ile Met Glu Arg Leu Thr Leu Arg Arg Ile Asp Pro Val Thr Gly
                185                 190                 195
Glu Arg Tyr His Leu Met Tyr Lys Pro Pro Pro Thr Met Glu Ile
                200                 205                 210
Gln Ala Arg Leu Leu Gln Asn Pro Lys Asp Ala Glu Glu Gln Val
                215                 220                 225
Lys Leu Lys Met Asp Leu Phe Tyr Arg Asn Ser Ala Asp Leu Glu
                230                 235                 240
Gln Leu Tyr Gly Ser Ala Ile Thr Leu Asn Gly Asp Gln Asp Pro
                245                 250                 255
Tyr Thr Val Phe Glu Tyr Ile Glu Ser Gly Ile Ile Asn Pro Leu
                260                 265                 270
Pro Lys Lys Ile Pro
                275
```

```
<210> 26
<211> 660
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480774CD1

<400> 26
Met Arg Leu Glu Ala Pro Arg Gly Gly Arg Arg Arg Gln Pro Gly
  1               5                  10                  15
Gln Gln Arg Pro Gly Pro Gly Ala Gly Ala Pro Ala Gly Arg Pro
                 20                  25                  30
Glu Gly Gly Gly Pro Trp Ala Arg Thr Glu Glu Ser Ser Leu His
                 35                  40                  45
Ser Glu Pro Glu Arg Ala Gly Leu Gly Pro Ala Pro Gly Thr Glu
                 50                  55                  60
Ser Pro Gln Ala Glu Phe Trp Thr Asp Gly Gln Thr Glu Pro Ala
                 65                  70                  75
Ala Ala Gly Leu Gly Val Glu Thr Glu Arg Pro Lys Gln Lys Thr
                 80                  85                  90
Glu Pro Asp Arg Ser Ser Leu Arg Thr His Leu Glu Trp Ser Trp
                 95                 100                 105
Ser Glu Leu Glu Thr Thr Cys Leu Trp Thr Glu Thr Gly Thr Asp
                110                 115                 120
Gly Leu Trp Thr Asp Pro His Arg Ser Asp Leu Gln Phe Gln Pro
                125                 130                 135
Glu Glu Ala Ser Pro Trp Thr Gln Pro Gly Val His Gly Pro Trp
                140                 145                 150
Thr Glu Leu Glu Thr His Gly Ser Gln Thr Gln Pro Glu Arg Val
                155                 160                 165
Lys Ser Trp Ala Asp Asn Leu Trp Thr His Gln Asn Ser Ser Ser
                170                 175                 180
Leu Gln Thr His Pro Glu Gly Ala Cys Pro Ser Lys Glu Pro Ser
                185                 190                 195
Ala Asp Gly Ser Trp Lys Glu Leu Tyr Thr Asp Gly Ser Arg Thr
                200                 205                 210
Gln Gln Asp Ile Glu Gly Pro Trp Thr Glu Pro Tyr Thr Asp Gly
                215                 220                 225
Ser Gln Lys Lys Gln Asp Thr Glu Ala Ala Arg Lys Gln Pro Gly
```

```
                    230                          235                          240
Thr Gly Gly Phe Gln Ile Gln Gln Asp Thr Asp Gly Ser Trp Thr
                    245                          250                          255
Gln Pro Ser Thr Asp Gly Ser Gln Thr Ala Pro Gly Thr Asp Cys
                    260                          265                          270
Leu Leu Gly Glu Pro Glu Asp Gly Pro Leu Glu Glu Pro Glu Pro
                    275                          280                          285
Gly Glu Leu Leu Thr His Leu Tyr Ser His Leu Lys Cys Ser Pro
                    290                          295                          300
Leu Cys Pro Val Pro Arg Leu Ile Ile Thr Pro Glu Thr Pro Glu
                    305                          310                          315
Pro Glu Ala Gln Pro Val Gly Pro Pro Ser Arg Val Glu Gly Gly
                    320                          325                          330
Ser Gly Gly Phe Ser Ser Ala Ser Ser Phe Asp Glu Ser Glu Asp
                    335                          340                          345
Asp Val Val Ala Gly Gly Gly Gly Ala Ser Asp Pro Glu Asp Arg
                    350                          355                          360
Ser Gly Ser Lys Pro Trp Lys Lys Leu Lys Thr Val Leu Lys Tyr
                    365                          370                          375
Ser Pro Phe Val Val Ser Phe Arg Lys His Tyr Pro Trp Val Gln
                    380                          385                          390
Leu Ser Gly His Ala Gly Asn Phe Gln Ala Gly Glu Asp Gly Arg
                    395                          400                          405
Ile Leu Lys Arg Phe Cys Gln Cys Glu Gln Arg Ser Leu Glu Gln
                    410                          415                          420
Leu Met Lys Asp Pro Leu Arg Pro Phe Val Pro Ala Tyr Tyr Gly
                    425                          430                          435
Met Val Leu Gln Asp Gly Gln Thr Phe Asn Gln Met Glu Asp Leu
                    440                          445                          450
Leu Ala Asp Phe Glu Gly Pro Ser Ile Met Asp Cys Lys Met Gly
                    455                          460                          465
Ser Arg Thr Tyr Leu Glu Glu Glu Leu Val Lys Ala Arg Glu Arg
                    470                          475                          480
Pro Arg Pro Arg Lys Asp Met Tyr Glu Lys Met Val Ala Val Asp
                    485                          490                          495
Pro Gly Ala Pro Thr Pro Glu Glu His Ala Gln Gly Ala Val Thr
                    500                          505                          510
Lys Pro Arg Tyr Met Gln Trp Arg Glu Thr Met Ser Ser Thr Ser
                    515                          520                          525
Thr Leu Gly Phe Arg Ile Glu Gly Ile Lys Lys Ala Asp Gly Thr
                    530                          535                          540
Cys Asn Thr Asn Phe Lys Lys Thr Gln Ala Leu Glu Gln Val Thr
                    545                          550                          555
Lys Val Leu Glu Asp Phe Val Asp Gly Asp His Val Ile Leu Gln
                    560                          565                          570
Lys Tyr Val Ala Cys Leu Glu Glu Leu Arg Glu Ala Leu Glu Ile
                    575                          580                          585
Ser Pro Phe Phe Lys Thr His Glu Val Val Gly Ser Ser Leu Leu
                    590                          595                          600
Phe Val His Asp His Thr Gly Leu Ala Lys Val Trp Met Ile Asp
                    605                          610                          615
Phe Gly Lys Thr Val Ala Leu Pro Asp His Gln Thr Leu Ser His
                    620                          625                          630
Arg Leu Pro Trp Ala Glu Gly Asn Arg Glu Asp Gly Tyr Leu Trp
                    635                          640                          645
Gly Leu Asp Asn Met Ile Cys Leu Leu Gln Gly Leu Ala Gln Ser
                    650                          655                          660
```

<210> 27
<211> 822
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2011384CB1

<400> 27
atgtcgggag acaaacttct gagcgaactc ggttataagc tgggccgcac aattggagag 60
ggcagctact ccaaggtgaa ggtggccaca tccaagaagt acaagggtac cgtggccatc 120
aaggtggtgg accggcggcg agcgcccccg gacttcgtca acaagttcct gccgcgagag 180
ctgtccatcc tgcggggcgt gcgacacccg cacatcgtgc acgtcttcga gttcatcgag 240
gtgtgcaacg ggaaactgta catcgtgatg gaagcggccg ccaccgacct gctgcaagcc 300
gtgcagcgca acgggcgcat ccccggagtt caggcgcgcg acctctttgc gcagatcgcc 360
ggcgccgtgc gctacctgca cgatcatcac ctggtgcacc gcgacctcaa gtgcgaaaac 420
gtgctgctga gcccggacga gcgccgcgtc aagctcaccg acttcggctt cggccgccag 480
gcccatggct acccagacct gagcaccacc tactgcggct cagccgccta cgcgtcaccc 540
gaggtgctcc tgggcatccc ctacgacccc aagaagtacg atgtgtggag catgggcgtc 600
gtgctctacg tcatggtcac cgggtgcatg cccttcgacg actcggacat cgccggcctg 660
ccccggcgcc agaaacgcgg cgtgctctat cccgaaggcc tcgagctgtc cgagcgctgc 720
aaggccctga tcgccgagct gctgcagttc agcccgtccg ccaggccctc cgcgggccag 780
gtagcgcgca actgctggct gcgcgccggg gactccggct ag 822

<210> 28
<211> 1376
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2004888CB1

<220>
<221> unsure
<222> 1369
<223> a, t, c, g, or other

<400> 28
gcttattgaa tatttaaata agagtcccag tgtggatcac ttgctatcca ttaagaagac 60
attgaaaagc ttaaaagctc tactcagatg gaaattggtt gaaaagagta atttggaaga 120
gtcagatgat cctgatggct ctcaaattga gaaaataaaa gaagaaataa ctcagctgcg 180
caataatgtc tttcaggaaa tttatcatga gagagaggaa tatgagatgc taactagttt 240
ggcacagaaa tggttccctg agctgcctct gcttcatcct gaaataggat tactcaaata 300
catgaactct ggtggtctcc ttacaatgag cttggaacga gatcttcttg atgctgagcc 360
catgaaggaa cttagcagca agcgtccttt ggtacgttct gaggttaatg ggcagataat 420
tctgttaaag ggctattctg tggatgttga cacagaagcc aaggtgattg agagagcagc 480
cacctaccat agagcttgga gagaagctga aggagactca gggttactgc cattgatatt 540
cctgttttta tgtaagtctg atcctatggc ttatctgatg gtcccatact accctagggc 600
aaacctgaat gctgttcaag ccaacatgcc tttaaattca gaagaaactt taaaggtcat 660
gaaaggtgtt gcccagggtc tgcatacatt gcataaggct gacataattc atggatcact 720
tcatcagaac aatgtatttg ctttaaaccg tgaacaagga attgttggag attttgactt 780
caccaaatct gtgagtcagc gagcctcggt gaacatgatg gttggtgact tgagtttgat 840
gtcacctgag ttgaaaatgg gaaaacctgc ttctccaggt tcagacttat atgcttatgg 900
ctgcctctta ttatggcttt ctgttcaaaa tcaggagttt gagataaata aagatggaat 960
ccccaaagtg gatcagtttc atctggatga taaagtcaaa tccctcctct gtagcttgat 1020
atgttataga agttcaatga ctgctgaaca agttttaaat gctgaatgtt tcttgatgcc 1080
aaaggagcaa tcagttccaa acccagaaaa agatactgaa tacaccctat ataaaaagga 1140
agaagaaata aagacggaga acttggataa atgtatggag aagacaagaa atggtgaagc 1200
caactttgat tgttaaatta ttattgttgt tgttgcagag gttcttttta aaaactttgg 1260
tttggttaat acacagaaat atctagaaat gttctgggac tagttgagtt gtatctttag 1320
tattcaggtt gaagaaaaat aaagatgtgt ggtatactag ttctgatgng ctgtgc 1376

<210> 29
<211> 3468
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2258952CB1

<400> 29
ttccactata acctttctct agggtcaaag agatgatgag tgacaccagc acgttcccca 60
atcacccttc ctcccctgct gcatccccat ctgggggaag gggagtcatg gccagccctg 120

```
cttgggacag gagcaaaggg tggtcccaga cccctcagag agctgacttt gtctctaccc 180
ccttgcaggt tcatactctc aggccagaga acctcctgct ggtgtccacc ttggatggaa 240
gtctccacgc actaagcaag cagacagggg acctgaagtg gactctgagg gatgatcccg 300
tcatcgaagg accaatgtac gtcacagaaa tggcctttct ctctgaccca gcagatggca 360
gcctgtacat cttggggacc caaaaacaac agggattaat gaaactgcca ttcaccatcc 420
ctgagctggt tcatgcctct ccctgccgca gctctgatgg ggtcttctac acaggccgga 480
agcaggatgc ctggtttgtg gtggaccctg agtcagggga gacccagatg acactgacca 540
cagagggtcc ctccaccccc cgcctctaca ttggccgaac acagtatacg gtcaccatgc 600
atgacccaag agccccagcc ctgcgctgga acaccaccta ccgccgctac tcagcgcccc 660
ccatggatgg ctcacctggg aaatacatga gccacctggc gtcctgcggg atgggcctgc 720
tgctcactgt ggacccagga agcgggacgg tgctgtggac acaggacctg ggcgtgcctg 780
tgatgggcgt ctacacctgg caccaggacg gcctgcgcca gctgccgcat ctcacgctgg 840
ctcgagacac tctgcatttc ctcgccctcc gctggggcca catccgactg cctgcctcag 900
gcccccggga cacagccacc ctcttctcta ccttggacac ccagctgcta atgacgctgt 960
atgtggggaa ggatgaaact ggcttctatg tctctaaagc actggtccac acaggagtgg 1020
ccctggtgcc tcgtggactg accctggccc ccgcagatgg ccccaccaca gatgaggtga 1080
cactccaagt ctcaggagag cgagagggct cacccagcac tgctgttaga taccccctcag 1140
gcagtgtggc cctcccaagc cagtggctgc tcattggaca ccacgagcta cccccagtcc 1200
tgcacaccac catgctgagg gtccatccca ccctggggag tggaactgca gagacaagac 1260
ctccagagaa tacccaggcc ccagccttct tcttggagct attgagcctg agccgagaga 1320
aactttggga ctccgagctg catccagaag aaaaaactcc agactcttac ttggggctgg 1380
gaccccaaga cctgctggca gctagcctca ctgctgtcct cctgggaggg tggattctct 1440
ttgtgatgag gcagcaacag gagaccccc tggcacctgc agactttgct cacatctccc 1500
aggatgccca gtccctgcac tcgggggcca gccggaggag ccagaagagg cttcagagtc 1560
cctcacctga gtcaccaccc tcctctcccc cagctgagca actcaccgta gtggggaaga 1620
tttccttcaa tcccaaggac gtgctgggcc gcggggcagg cgggactttc gttttcaggg 1680
gacagtttga gggacgggca gtggctgtca agcggctcct ccgcgagtgc tttggcctgg 1740
ttcggcggga agttcaactg ctgcaggagt ctgacaggca ccccaacgtg ctccgctact 1800
tctgcaccga gcggggaccc cagttccact acattgccct ggagctctgc cgggcctcct 1860
tgcaggagta cgtagaaaac ccggacctgg atcgcggggg tctggagccc gaggtcgtgc 1920
tgcagcagct gatgtctggc ctggcccacc tgcactcttt acacatagtg caccgggacc 1980
tgaagccagg aaatattctc atcaccgggc ctgacagcca gggcctgggc agagtggtgc 2040
tctcagactt cggcctctgc aagaagctgc ctgctggccg ctgtagcttc agcctccact 2100
ccggcatccc cggcacggaa ggctggatgg cgcccgagct tctgcagctc ctgccaccag 2160
acagtcctac cagcgctgtg gacatcttct ctgcaggctg cgtgttctac tacgtgcttt 2220
ctggtggcag ccacccccttt ggagacagtc tttatcgcca ggcaaacatc ctcacagggg 2280
ctccctgtct ggctcacctg gaggaagagg tccacgacaa ggtggttgcc cgggacctgg 2340
ttggagccat gttgagccca ctgccgcagc cacgcccctc tgccccccag gtgctggccc 2400
acccccttctt ttggagcaga gccaagcaac tccagttctt ccaggacgtc agtgactggc 2460
tggagaagga gtccgagcag gagcccctgg tgagggcact ggaggcggga ggctgcgcag 2520
tggtccggga caactggcac gagcacatct ccatgccgct gcagacagat ctgagaaagt 2580
tccggtccta taaggggaca tcagtgcgag acctgctccg tgctgtgagg aacaagaagc 2640
accactacag ggagctccca gttgaggtgc gacaggcact cggccaagtc cctgatggct 2700
tcgtccagta cttcacaaac cgcttcccac ggctgctcct ccacacgcac cgagccatga 2760
ggagctgcgc ctctgagagc ctcttcctgc cctactaccc gccagactca gaggccagga 2820
ggccatgccc tggggccaca gggaggtgag gtgggctgga tgccacacag atggtctccg 2880
tgctggctca ctgaagagct gagcctgtgg ctggcctcag aatcaggctg ggtgcagtgg 2940
ctcacacctg taatcccagc atttgggag gctgagtgag aggatcactt gagctcagga 3000
gttcgagacc agcctggcca acatggcaac accccatttc tacaaaaaat ttgtaaaatt 3060
agccaggcat ggtggcgcac gcctgtagtc ccagctgctt gggaggctga ggtgggagaa 3120
tcacttgagc ccaggagttc gaggctgcag tgagccagga tcatgccact gcactccagc 3180
ctggtccaca gagagacact gtcacccccct ttcccccaca agactggcag aggctgggca 3240
gcctggggct gatgaagcag agatgttcgc tggatcccag gccctggcac ccctcaggaa 3300
atacaagaaa aagaatattc acatctgttt aatgtgcata aagccaagga aaggacagtt 3360
ccgaattcaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3420
aaaaaaaaaa aaaaagaaga aaaaaaaaaa aaagaagaag acccagac 3468
```

<210> 30
<211> 2831
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473244CB1

<400> 30

cttctcccgc cggggccgct tgttgcaccg ccccgcggcc tgcgggagcc gctcgccccg 60
gccttgtgct cgcgtccgca ccccttttcct gtcgcccccc ggggcccgca ccacagcccg 120
gccggcgaga ccccggccag accccgctgc ccgcacaaaa tgtcggcccg gacgccattg 180
ccgacggtga acgagcggga cacggaaaat catacatctg tggatggata tactgaacca 240
cacatccagc ctaccaagtc gagtagcaga cagaacatcc cccggtgtag aaactccatt 300
acgtcagcaa cagatgaaca gcctcacatt ggaaattacc gtttacaaaa aacaataggg 360
aagggaaatt ttgccaaagt caaattggca agacacgttc taactggtag agaggttgct 420
gtgaaaataa tagacaaaac tcagctaaat cctaccagtc tacaaaagtt atttcgagaa 480
gtacgaataa tgaagatact gaatcatcct aatatagtaa aattgtttga agttattgaa 540
acagagaaga ctctctattt agtcatggaa tacgcgagtg gggtgaagt atttgattac 600
ttagttgccc atggaagaat gaaagagaaa gaggcccgtg caaaatttag gcagattgta 660
tctgctgtac agtattgtca tcaaaagtac attgttcacc gtgatcttaa ggctgaaaac 720
cttctccttg atggtgatat gaatattaaa attgctgact ttggttttag taatgaattt 780
acagttggga acaaattgga cacattttgt ggaagcccac cctatgctgc tcccgagctt 840
ttccaaggaa agaagtatga tgggcctgaa gtggatgtgt ggagtctggg cgtcattctc 900
tatacattag tcagtggctc cttgcctttc gatggccaga atttaaagga actgcgagag 960
cgagttttac gagggaagta ccgtattccc ttctatatgt ccacagactg tgaaaatctt 1020
ctgaagaaat tattagtcct gaatccaata aagagaggca gcttggaaca aataatgaaa 1080
gatcgatgga tgaatgttgg tcatgaagag gaagaactaa agccatatac tgagcctgat 1140
ccggatttca atgacacaaa aagaatagac attatggtca ccatgggctt tgcacgagat 1200
gaaataaatg atgccttaat aaatcagaag tatgatgaag ttatggctac ttatattctt 1260
ctaggtagaa aaccacctga atttgaaggt ggtgaatcgt tatccagtgg aaacttgtgt 1320
cagaggtccc ggcccagtag tgacttaaac aacagcactc ttcagtcccc tgctcacctg 1380
aaggtccaga gaagtatctc agcaaatcag aagcagcggc gtttcagtga tcatgctggt 1440
ccatccattc ctcctgctgt atcatatacc aaaagacctc aggctaacag tgtggaaagt 1500
gaacagaaag aggagtggga caaagatgtg gctcgaaaac ttggcagcac aacagttgga 1560
tcaaaaagcg agatgactgc aagccctctt gtagggccag agaggaaaaa atcttcaact 1620
attccaagta acaatgtgta ttctggaggt agcatggcaa gaaggaatac atatgtctgt 1680
gaaaggacca cagatcgata cgtagcattg cagaatggaa aagacagcag ccttacggag 1740
atgtctgtga gtagcatatc ttctgcaggc tcttctgtgg cctctgctgt cccctcagca 1800
cgaccccgcc accagaagtc catgtccact tctggtcatc ctattaaagt cacactgcca 1860
accattaaag acggctctga agcttaccgg cctggtacaa cccagagagt gcctgctgct 1920
tccccatctg ctcacagtat tagtactgcg actccagacc ggacccgttt tccccgaggg 1980
agctcaagcc gaagcacttt ccatggtgaa cagctccggg agcgacgcag cgttgcttat 2040
aatgggccac ctgcttcacc atcccatgaa acgggtgcat ttgcacatgc cagaaggggа 2100
acgtcaactg gtataataag caaaatcaca tccaaatttg ttcgcaggga tccaagtgaa 2160
ggcgaagcca gtggcagaac cgacacctca agaagtacat caggggaacc aaaagaaaga 2220
gacaaggaag agggtaaaga ttctaagccg cgttctttgc ggttcacatg gagtatgaag 2280
accactagtt caatggaccc taatgacatg atgagagaaa tccgaaaagt gttagatgca 2340
aataactgtg attatgagca aaaagagaga tttttgcttt tctgtgtcca tggagacgct 2400
agacaggata gcctcgtgca gtgggagatg gaagtctgca agttgccacg actgtcactt 2460
aatggggttc gcttcaagcg aatatctggg acatctattg cctttaagaa cattgcatca 2520
aaaatagcaa atgagcttaa gctgtaaaga agtccaaatt tacaggttca gggaagatac 2580
atacatatat gaggtacagt ttttgaatgt actggtaatg cctaatgtgg tctgcctgtg 2640
aatctcccca tgtagaattt gcccttaatg caataaggtt atacatagtt atgaactgta 2700
aaattaaagt cagtatgaac tataataaat atctgtagct taaaaagtag gttcacatgt 2760
acaggtaagt atattgtgta tttctgttca ttttctgttc atagagttgt ataataaaac 2820
atgattgctt t 2831

<210> 31
<211> 2693
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1242491CB1

<400> 31
agtgtgctgg aaagattgcc cctgacttga tttggctgac ctgcctagaa atattatgtt 60
gaataatgat gagttggaat ttgaacaagc tccagagttt tctcctaggt gatggcagtt 120
ttggatcagt ttaccgagca gcctatgaag gagaagaagt ggctgtgaag atttttaata 180
aacatacatc actcaggctg ttaagacaag agcttgtggt gctttgccac ctccaccacc 240
ccagtttgat atctttgctg gcagctggga ttcgtccccg gatgttggtg atggagttag 300
cctccaaggg ttccttggat cgcctgcttc agcaggacaa agccagcctc actagaaccc 360
tacagcacag gattgcactc cacgtagctg atggtttgag atacctccac tcagccatga 420
ttatataccg agacctgaaa ccccacaatg tgctgctttt cacactgtat cccaatgctg 480

```
ccatcattgc aaagattgct gactacggca ttgctcagta ctgctgtaga atggggataa 540
aaacatcaga gggcacacca gggtttcgtg cacctgaagt tgccagagga aatgtcattt 600
ataaccaaca ggctgatgtt tattcatttg gtttactact ctatgacatt ttgacaactg 660
gaggtagaat agtagagggt ttgaagtttc caaatgagtt tgatgaatta gaaatacaag 720
gaaaattacc tgatccagtt aaagaatatg gttgtgcccc atggcctatg gttgagaaat 780
taattaaaca gtgtttgaaa gaaaatcctc aagaaaggcc tacttctgcc caggtctttg 840
acattttgaa ttcagctgaa ttagtctgtc tgacgagacg cattttatta cctaaaaacg 900
taattgttga atgcatggtt gctacacatc acaacagcag gaatgcaagc atttggctgg 960
gctgtgggca caccgacaga ggacagctct catttcttga cttaaatact gaaggataca 1020
cttctgagga agttgctgat agtagaatat tgtgcttagc cttggtgcat cttcctgttg 1080
aaaaggaaag ctggattgtg tctgggacac agtctggtac tctcctggtc atcaataccg 1140
aagatgggaa aaagagacat accctagaaa agatgactga ttctgtcact tgtttgtatt 1200
gcaattcctt ttccaagcaa agcaaacaaa aaaatttttct tttggttgga accgctgatg 1260
gcaagttagc aatttttgaa gataagactg ttaagcttaa aggagctgct cctttgaaga 1320
tactaaatat aggaaatgtc agtactccat tgatgtgttt gagtgaatcc acaaattcaa 1380
cggaaagaaa tgtaatgtgg ggaggatgtg gcacaaagat tttctccttt tctaatgatt 1440
tcaccattca gaaactcatt gagacaagaa caagccaact gttttcttat gcagctttca 1500
gtgattccaa catcataaca gtggtggtag acactgctct ctatattgct aagcaaaata 1560
gcccctgttgt ggaagtgtgg gataagaaaa ctgaaaaact ctgtggacta atagactgcg 1620
tgcacttttt aagggaggta acggtaaaag aaaacaagga atcaaaacac aaaatgtctt 1680
attctgggag agtgaaaacc ctctgccttc agaagaacac tgctctttgg ataggaactg 1740
gaggaggcca tattttactc ctggatcttt caactcgtcg acttatacgt gtaatttaca 1800
acttttgtaa ttcggtcaga gtcatgatga cagcacagct aggaagcctt aaaaatgtca 1860
tgctggtatt gggctacaac cggaaaaata ctgaaggtac acaaaagcag aaagagatac 1920
aatcttgctt gaccgtttgg gacatcaatc ttccacatga agtgcaaaat ttagaaaaac 1980
acattgaagt gagaaaagaa ttagctgaaa aaatgagacg aacatctgtt gagtaagaga 2040
gaaataggaa ttgtctttgg ataggaaaat tattctctcc tcttgtaaat atttatttta 2100
aaaatgttca catggaaagg gtactcacat tttttgaaat agctcgtgtg tatgaaggaa 2160
tgttattatt tttaatttaa atatatgtaa aaatacttac cagtaaatgt gtattttaaa 2220
gaactattta aaacacaatg ttatatttct tataaatacc agttactttc gttcattaat 2280
taatgaaaat aaatctgtga agtacctaat ttaagtactc atactaaaat ttataaggcc 2340
gataattttt tgttttcttg tctgtaatgg aggtaaactt tattttaaat tctgtgctta 2400
agacaggact attgcttgtc gatttttcta gaaatctgca cggtataatg aaaatattaa 2460
gacagtttcc catgtaatgt attccttctt agattgcatc gaaatgcact atcatatatg 2520
cttgtaaata ttcaaatgaa tttgcactaa taaagtcctt tgttggtatg tgaattctct 2580
ttgttgctgt tgcaaacagt gcatcttaca caacttcact caattcaaaa gaaaactcca 2640
ttaaaagtac taatgaaaaa acatgacata ctgtcaaagt cctcatatct agg        2693

<210> 32
<211> 2973
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2634875CB1

<400> 32
agtgtgctgg aaagactgcc cacaccccctg cctccgcctc tgcccacccg gcccaatccc 60
ttacaactgc ccaggactgc tcctgagcag ccgctgggag acagacggca accaggttgc 120
ccctctttgc tccaggtacc tctctcccct cagttagcag gcctcggctt cctgtctcac 180
tgcagccaga cgagagggga aattggacag cctgacacac tccactcttg tttctgcagc 240
tagaaagact tgagttagac aagcagcagc acacgcctcc ctacctcatg gcgacagaaa 300
atggagcagt tgagctggga attcagaacc catcaacaga caaggcacct aaaggtccca 360
caggtgaaag acccctggct gcagggaaag accctggccc cccagaccca aagaaagctc 420
cggatccacc caccctgaag aaagatgcca aagcccctgc ctcagagaaa ggggatggta 480
ccctggccca accctcaact agcagccaag gccccaaagg agagggtgac aggggcgggg 540
ggcccgcgga gggcagtgct gggcccccgg cagccctgcc ccagcagact gcgacacctg 600
agaccagcgt caagaagccc aaggctgagc agggagcctc aggcagccag gatcctggaa 660
agcccagggt gggcaagaag gcagcagagg gccaagcagc agccaggagg ggctcacctg 720
cctttctgca tagccccagc tgtcctgcca tcatctccag ttctgagaag ctgctggcca 780
agaagccccc aagcgaggca tcagagctca cctttgaagg ggtgcccatg acccacagcc 840
ccacggatcc caggccagcc aaggcagaag aaggaaagaa catcctggca gagagccaga 900
aggaagtggg agagaaaacc ccaggccagg ctggccaggc taagatgcaa ggggacacct 960
cgagggggat tgagttccag gctgttccct cagagaaatc cgaggtgggg caggccctct 1020
gtctcacagc cagggaggag gactgcttcc agattttgga tgattgcccg ccacctccgg 1080
ccccttccc tcaccgcatg gtggagctga ggaccgggaa tgtcagcagt gaattcagta 1140
```

tgaactccaa ggaggcgctc ggaggtggca agtttggggc agtctgtacc tgcatggaga 1200
aagccacagg cctcaagctg gcagccaagg tcatcaagaa acagactccc aaagacaagg 1260
aaatggtgtt gctggagatt gaggtcatga accagctgaa ccaccgcaat ctgatccagc 1320
tgtatgcagc catcgagact ccgcatgaga tcgtcctgtt catggagtac atcgagggcg 1380
gagagctctt cgagaggatt gtggatgagg actaccatct gaccgaggtg gacaccatgg 1440
tgtttgtcag gcagatctgt gacgggatcc tcttcagtgt gctggaaagg gttttgcacc 1500
tggacctcaa gccagagaac atcctgtgtg tcaacaccac cgggcatttg gtgaagatca 1560
ttgactttgg cctggcacgg aggtataacc ccaacgagaa gctgaaggtg aactttggga 1620
ccccagagtt cctgtcacct gaggtggtga agggtgacca aatctccgat aagacagaca 1680
tgtggagtat gggggtgatc acctacatgc tgctgagcgg cctctccccc ttcctgggag 1740
atgatgacac agagaccct aacaacgttc tatctggcaa ctggtacttt gatgaagaga 1800
cctttgaggc cgtatcagac gaggccaaag actttgtctc caacctcatc gtcaaggacc 1860
agagggcccg gatgaacgct gcccagtgtc tcgcccatcc ctggctcaac aacctggcgg 1920
agaaagccaa acgctgtaac cgacgcctta agtcccagat cttgcttaag aaatacctca 1980
tgaagaggcg ctggaagaaa aacttcattg ctgtcagcgc tgccaaccgc ttcaagaaga 2040
tcagcagctc gggggcactg atggctctgg gggtctgagc cctgggcgca gctgaagcct 2100
ggacgcagcc acacagtggc cggggctgaa gccacacagc ccagaaggcc agaaaaggca 2160
gccagatccc cagggcagcc tcgttaggac aaggctgtgc caggctggga ggctcggggc 2220
tccccacgcc cccatgcagt gaccgcttcc ccgatgtgag ccgcctcgga gtgtggcctg 2280
gatccatcct gctagcacct ccccagacag ggctccagcc tgtcggccac accccagact 2340
ccaggccccc gttgaagccg ctcccggttc cctccccagc tcctcgtctt tgaactgccg 2400
ccgccgtggt gacccctgct ttgccccact gggagagtcc ttagcctggg cctcctccta 2460
gctggagtgc catggctggg ggtctcagc atgtagggct tctgtggttg tggatgggag 2520
gctcctggtg gggcagaaag gctgcaacgc tgattcctaa ggcccagctg ccagggaaga 2580
cagagcaggc tttgtgagag aggacctcca tgcccccgcc acctccccac tccagcagat 2640
aaggccgagc ccacaccatc tggcccaggc tggcccccac ccaccttcct tgcgaccacc 2700
aacacacagg aactctgtgt gagagagagg gcgcccagcc caggcctggt ggaggggggag 2760
gggagaagcc aagggacaca ggagaccacc cccgagcttg cctcagggcc aagccggccc 2820
aacccaacca ctcggggccc ccatcttggg ggtcacccat ggcctcagat gatggggtca 2880
gcaggcccag gagaattagg aaggccatgg ggcagcctcc agtctgctct cagcttgtgc 2940
cttgtaaata aatgtacagg ttggaaaaaa aaa 2973

<210> 33
<211> 2066
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3951059CB1

<400> 33
cgccagtggg gagatgttga agttcaaata tggagcgcgg aatcctttgg atgctggtgc 60
tgctgaaccc attgccagcc gggcctccag gctgaatctg ttcttccagg ggaaaccacc 120
ctttatgact caacagcaga tgtctcctct ttcccgagaa gggatattag atgccctctt 180
tgttctcttt gaagaatgca gtcagcctgc tctgatgaag attaagcacg tgagcaactt 240
tgtccggaag tattccgaca ccatagctga gttacaggag ctccagcctt cggcaaagga 300
cttcgaagtc agaagtcttg taggttgtgg tcactttgct gaagtgcagg tggtaagaga 360
gaaagcaacc ggggacatct atgctatgaa agtgatgaag aagaaggctt tattggccca 420
ggagcaggtt tcattttttg aggaagagcg gaacatatta tctcgaagca caagcccgtg 480
gatcccccaa ttacagtatg cctttcagga caaaaatcac ctttatctgg tcatggaata 540
tcagcctgga ggggacttgc tgtcactttt gaatagatat gaggaccagt tagatgaaaa 600
cctgatacag ttttacctag ctgagctgat tttggctgtt cacagcgttc atctgatggg 660
atacgtgcat cgagacatca agcctgagaa cattctcgtt gaccgcacag gacacatcaa 720
gctggtggat tttggatctg ccgcgaaaat gaattcaaac aagatggtga atgccaaact 780
cccgattggg accccagatt acatggctcc tgaagtgctg actgtgatga acggggatgg 840
aaaaggcacc taccgcctgg actgtgactg gtggtcagtg ggcgtgattg cctatgagat 900
gatttatggg agatccccct tcgcagaggg aacctctgcc agaaccttca ataacattat 960
gaatttccag cggtttttga aatttccaga tgaccccaaa gtgagcagtg actttcttga 1020
tctgattcaa agcttgttgt gcggccagaa agagagactg aagtttgaag gtctttgctg 1080
ccatcctttc ttctctaaaa ttgactggaa caacattcgt aactctcctc ccccttcgt 1140
tcccaccctc aagtctgacg atgacacctc caattttgat gaaccagaga agaattcgtg 1200
ggtttcatcc tctccgtgcc agctgagccc ctcaggcttc tcgggtgaag aactgccgtt 1260
tgtggggttt tcgtacagca aggcactggg gattcttggt agatctgagt ctgttgtgtc 1320
gggtctggac tcccctgcca agactagctc catggaaaag aaacttctca tcaaaagcaa 1380
agagctacaa gactctcagg acaagtgtca caaggtattt atttccgcag ccggcctcct 1440
tccttgctcc aggatcctcc cgtccgtata tgccaaggga tccgcccggg gccgctgctg 1500

gctctgagcc gcctgatccg tagagagtga ggcgctcctg ccttcgctga agtcgcgcct 1560
ccagcagctc agagggagat gaattcgggc cttgctgttg ctgtaaatcc tttaaatcta 1620
aaccagagga ggccctggat ttaaacagtc cgtttctcag catgacccag ccagatgtct 1680
gcttcttccg gcaggtggcc tgggtcctca cctgtggctg agatacatcc catctgcttt 1740
gagtgatgcg aagtctctct tcctagtctt ttaaaactcc tgcttatgtc actgcggcca 1800
ctgtgttgat tacgctcaac gtctcttaac attcactgtt cctgcccaga ggcaacgctc 1860
tggaaactaa taagtcactg cttgcctggg actcctaaga gtgcagacga ataaatatct 1920
ccttgccctg tcctggattt gtcctctaga tctttgcaag gagatggggg gggatcaaga 1980
tggatttggg ataaaattaa agtgacgtct gcaaaaacaa aacaaaaaca aaagcaaaca 2040
ggtgaaaaat gatgattgtg gcttcc 2066

<210> 34
<211> 3975
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7395890CB1

<400> 34
agtgtgctgg aaagggcggc ctcggctgcg ccgagagcgg agacacaggc tcaagatggc 60
agattccgac tgaggctggg ggggccgagc tcgcgcgccg ctttcccgtc cccgttgcca 120
tgaaccgcgg acaccccggc cccgatggcc cccgtgtacg aaggtatggc ctcacatgtg 180
caagttttct cccctcacac ccttcaatca agtgccttct gtagtgtgaa gaaactgaaa 240
atagagccga gttccaactg ggacatgact gggtacggct cccacagcaa agtgtatagc 300
cagagcaaga acatccccct gtcgcagcca gccaccacaa ccgtcagcac ctccttgccg 360
gtcccaaacc caagcctacc ttacgagcag accatcgtct tcccaggaag caccgggcac 420
atcgtggtca cctcagcaag cagcacttct gtcaccgggc aagtcctcgg cggaccacac 480
aacctaatgc gtcgaagcac tgtgagcctc cttgatacct accaaaaatg tggactcaag 540
cgtaagagcg aggagatcga gaacacaagc agcgtgcaga tcatcgagga gcatccaccc 600
atgattcaga ataatgcaag cggggccact gtcgccactg ccaccacgtc tactgccacc 660
tccaaaaaca gcggctccaa cagcgagggc gactatcagc tggtgcagca tgaggtgctg 720
tgctccatga ccaacaccta cgaggtctta gagttcttgg gccgagggac gtttgggcaa 780
gtggtcaagt gctggaaacg gggcaccaat gagatcgtag ccatcaagat cctgaagaac 840
cacccatcct atgcccgaca aggtcagatt gaagtgagca tcctggcccg gttgagcacg 900
gagagtgccg atgactataa cttcgtccgg gcctacgaat gcttccagca caagaaccac 960
acgtgcttgg tcttcgagat gttggagcag aacctctatg actttctgaa gcaaaacaag 1020
tttagcccct tgcccctcaa atacattcgc ccagttctcc agcaggtagc cacagccctg 1080
atgaaactca aaagcctagg tcttatccac gctgacctca aaccagaaaa catcatgctg 1140
gtggatccat ctagacaacc atacagagtc aaggtcatcg actttggttc agccagccac 1200
gtctccaagg ctgtgtgctc cacctacttg cagtccagat attacagggc ccctgagatc 1260
atccttggtt taccattttg tgaggcaatt gacatgtggt ccctgggctg tgttattgca 1320
gaattgttcc tgggttggcc gttatatcca ggagcttcgg agtatgatca gattcggtat 1380
atttcacaaa cacagggttt gcctgctgaa tatttattaa gcgccgggac aaagacaact 1440
aggtttttca accgtgacac ggactcacca tatcctttgt ggagactgaa gacaccagat 1500
gaccatgaag cagagacagg gattaagtca aaagaagcaa gaaagtacat tttcaactgt 1560
ttagatgata tggcccaggt gaacatgacg acagatttgg aagggagcga catgttggta 1620
gaaaaggctg accggcggga gttcattgac ctgttgaaga agatgctgac cattgatgct 1680
gacaagagaa tcactccaat cgaaaccctg aaccatccct ttgtcaccat gacacactta 1740
ctcgattttc cccacagcac acacgtcaaa tcatgtttcc agaacatgga gatctgcaag 1800
cgtcgggtga atatgtatga cacggtgaac cagagcaaaa ccccttttcat cacgcacgtg 1860
gcccccagca cgtccaccaa cctgaccatg acctttaaca accagctgac cactgtccac 1920
aaccagccct cagcggcatc catggctgca gtggcccagc ggagcatgcc cctgcagaca 1980
ggaacagccc agatttgtgc ccggcctgac ccgttccagc aagctctcat cgtgtgtccc 2040
cccggcttcc aaggcttgca ggcctctccc tctaagcacg ctggctactc ggtgcgaatg 2100
gaaaatgcag ttcccatcgt cactcaagcc ccaggagctc agcctcttca gatccaacca 2160
ggtctgcttg cccagcaggc ttggccaagt gggacccagc agatcctgct tcccccagca 2220
tggcagcaac tgactggagt ggccacccac acatcagtgc agcatgccac cgtgattccc 2280
gagaccatgg caggcaccca gcagctggcg gactggagaa atacgcatgc tcacggaagc 2340
cattataatc ccatcatgca gcagcctgca ctattgaccg gtcatgtgac ccttccagca 2400
gcacagccct taaatgtggg tgtggcccac gtgatgcggc agcagccaac cagcaccacc 2460
tcctcccgga agagtaagca gcaccagtca tctgtgagaa atgtctccac ctgtgaggtg 2520
tcctcctctc aggccatcag ctccccacag cgatccaagc gtgtcaagga gaacacacct 2580
ccccgctgtg ccatggtgca cagtagcccg gcctgcagca cctcggtcac ctgtgggtgg 2640
ggcgacgtgg cctccagcac cacccgggaa cggcagcggc agacaattgt cattcccgac 2700
actcccagcc ccacggtcag cgtcatcacc atcagcagtg acacggacga ggaggaggaa 2760

cagaaacacg cccccaccag cactgtctcc aagcaaagaa aaaacgtcat cagctgtgtc 2820
acagtccacg actcccccta ctccgactcc tccagcaaca ccagccccta ctccgtgcag 2880
cagcgtgctg ggcacaacaa tgccaatgcc tttgacacca aggggagcct ggagaatcac 2940
tgcacgggga accccсgaac catcatcgtg ccacccctga aaacccaggc cagcgaagta 3000
ttggtggagt gtgatagcct ggtgccagtc aacaccagtc accactcgtc ctcctacaag 3060
tccaagtcct ccagcaacgt gacctccacc agcggtcact cttcagggag ctcatctgga 3120
gccatcacct accggcagca gcggccgggc ccccacttcc agcagcagca gccactcaat 3180
ctcagccagg ctcagcagca catcaccacg gaccgcactg ggagccaccg aaggcagcag 3240
gcctacatca ctcccaccat ggcccaggct ccgtactcct tcccgcacaa cagccccagc 3300
cacggcactg tgcacccgca tctggctgca gccgctgccg ctgcccacct ccccacccag 3360
ccccacctct acacctacac tgcgccggcg gccctgggct ccaccggcac cgtggcccac 3420
ctggtggcct cgcaaggctc tgcgcgccac accgtgcagc acactgccta cccagccagc 3480
atcgtccacc aggtccccgt gagcatgggc cccсgggtcc tgccctcgcc caccatccac 3540
ccgagtcagt atccagccca atttgcccac cagacctaca tcagcgcctc gccagcctcc 3600
accgtctaca ctggataccc actgagcccc gccaaggtca accagtaccc ttacatataa 3660
acactggagg ggagggaggg agggagggag ggagagaatg gcccgaggga ggagggagag 3720
aaggagggag gcgctcctgg gaccgtgggc gctggccttt tatactgaag atgccgcaca 3780
caaacaatgc aaacggggca ggtgcggggg gggggggggc agagggcagg ggcacggggt 3840
cgggacacca gtgaaacttg aaccgggaag tgggaggacg tagagcagag aagagaacat 3900
ttttaaaagg aagggattaa agagggtggg aaatctatgg tttttatttt aaaaaagaaa 3960
aaggaaaaaa aaaaa 3975

<210> 35
<211> 1918
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7475546CB1

<400> 35
ccgcccgcag cgaggaagcg cccgcgcggg cgcaggcggc cgggatggcg gggcccggct 60
ggggtccccc gcgcctggac ggcttcatcc tcaccgagcg cctgggcagc ggcacgtacg 120
ccacggtgta caaggcctac gccaagaagg acactcgtga agtggtagcc ataaagtgtg 180
tagccaagaa aagtctgaac aaggcatcgg tggagaacct cctcacggag attgagatcc 240
tcaagggcat tcgacatccc cacattgtgc agctgaaaga ctttcagtgg gacagtgaca 300
atatctacct catcatggag ttttgcgcag gggcgacct gtctcgcttc atccataccc 360
gcaggattct gcctgagaag gtggcgcgtg tcttcatgca gcaattagct agcgccctgc 420
aattcctgca tgaacggaat atctctcacc tggatctgaa gccacagaac attctactga 480
gctccttgga gaagccccac ctaaaactgg cagactttgg tttcgcacaa cacatgtccc 540
cgtgggatga gaagcacgtg ctccgtggct cccccctcta catggccccc gagatggtgt 600
gccagcggca gtatgacgcc cgcgtggacc tctggtccat gggggtcatc ctgtatgaag 660
ccctcttcgg gcagcccccc tttgcctcca ggtcgttctc ggagctggaa gagaagatcc 720
gtagcaaccg ggtcatcgag ctccccttgc ggcccctgct ctcccgagac tgccgggacc 780
tactgcagcg gctcctggag cgggacccca gccgtcgcat ctccttccag gacttctttg 840
cgcacccctg ggtggacctg gagcacatgc ccagtgggga gagtctgggg cgagcaaccg 900
ccctggtggt gcaggctgtg aagaaagacc aggaggggga ttcagcagcc gccttatcac 960
tctactgcaa ggctctggac ttctttgtac ctgccctgca ctatgaagtg gatgcccagc 1020
ggaaggaggc aattaaggca aaggtggggc agtacgtgtc ccgggctgag gagctcaagg 1080
ccatcgtctc ctcttccaat caggccctgc tgaggcaggg gacctctgcc cgagacctgc 1140
tcagagagat ggcccgggac aagccacgcc tcctagctgc cctggaagtg gcttcagctg 1200
ccatggccaa ggaggaggcc gccggcgggg agcaggatgc cctggacctg taccagcaca 1260
gcctggggga gctactgctg ttgctggcag cggagccccc gggccggagg cgggagctgc 1320
ttcacactga ggttcagaac ctcatggccc gagctgaata cttgaaggag cagatgaggg 1380
aatctcgctg ggaagctgac accctggaca aagagggact gtcggaatct gttcgtagct 1440
cttgcaccct tcagtgaccc tagaagaatg attggacaga tgtgagccat ctggagcaga 1500
ggggcactaa cccaggctga cgccaagaat gaagtggccc actgcagccc tggcgagcag 1560
gcttcttgga tggacagtgc tgagaccccc atatcccaga gtccccagcc tccctcaggt 1620
tactctgcac cccacagatg gtttgatggc tgtgctgtat actggagggg agggcaggac 1680
tctgggagaa cagcacttct ttcatgagac ctttgttact cggtggttac tgggtcctgt 1740
gcctgtccgt tttggggcat gcagccctct atcatttttg gctccgagaa gagggcaagg 1800
ggcccccgca gggtacttct gtgcttgccc tcgccctgcc agcaggcagc tgtgcccctg 1860
gcctggcctt cccgggaccc cttattccaa ctcagctcct ctttgcactg gaatgggg 1918

<210> 36
<211> 1689

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477076CB1

<400> 36
agtgtgctgg aaagctttcc agacccctcc ctcccgctcc tgggaaagag agaaaccacc 60
gctgcgggtg ggtagagaag cacttggcgc ctcggggagg ggaccgcgcc cgcctcattt 120
gcgccttgca gcactgctgg accaggttac aagatgttca cctaagattg agacctagtg 180
actacatttc ctacgggaac aaataaatgg tttttcatct cccggagata cattacaaac 240
aaatatggtg ctaaaagaac tccttacctt tctctgacta caatttattt ggacatactt 300
ttgtattgaa gagaggtata catactgaag ctacttgctg tactatagga gactctgtcc 360
tgtaggatca tggaccatcc tagtagggaa aaggatgaaa gacaacggac aactaaaccc 420
atggcacaaa ggagtgcaca ctgctctcga ccatctggct cctcatcgtc ctctggggtt 480
cttatggtgg gacccaactt cagggttggc aagaagatag gatgtgggaa cttcggagag 540
ctcagattag gtaaaaatct ctacaccaat gaatatgtag caatcaaact ggaaccaata 600
aaatcacgtg ctccacagct tcatttagag tacagatttt ataaacagct tggcagtgca 660
ggtgaaggtc tcccacaggt gtattacttt ggaccatgtg ggaaatataa tgccatggtg 720
ctggagctcc ttggccctag cttggaggac ttgtttgacc tctgtgaccg aacatttact 780
ttgaagacgg tgttaatgat agccatccag ctgctttctc gaatggaata cgtgcactca 840
aagaacctca tttaccgaga tgtcaagcca gagaacttcc tgattggtcg acaaggcaat 900
aagaaagagc atgttataca cattatagac tttggactgg ccaaggaata cattgacccc 960
gaaaccaaaa aacacatacc ttatagggaa cacaaaagtt taactggaac tgcaagatat 1020
atgtctatca acacgcatct tggcaaagag caaagccgga gagatgattt ggaagcccta 1080
ggccatatgt tcatgtattt ccttcgaggc agcctcccct ggcaaggact caaggctgac 1140
acattaaaag agagatatca aaaaattggt gacaccaaaa ggaatactcc cattgaagct 1200
ctctgtgaga actttccaga ggagatggca acctaccttc gatatgtcag gcgactggac 1260
ttctttgaaa aacctgatta tgagtattta cggaccctct tcacagacct ctttgaaaag 1320
aaaggctaca cctttgacta tgcctatgat tgggttggga gacctattcc tactccagta 1380
gggtcagttc acgtagattc tggtgcatct gcaataactc gagaaagcca cacacatagg 1440
gatcggccat cacaacagca gcctcttcga aatcaggtgg ttagctcaac caatggagag 1500
ctgaatgttg atgatcccac gggagcccac tccaatgcac caatcacagc tcatgccgag 1560
gtggaggtag tggaggaagc taagtgctgc tgtttcttta agaggaaacg gaagaagact 1620
gctcagcgcc acaagtgacc agtgcctccc aggagtcctc agggcctggg ggactctgac 1680
tcaattgta 1689

<210> 37
<211> 1054
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1874092CB1

<400> 37
ggctggatgc tgcgatcccg caggtgagcg cagcaccctc cagccttgca gaagcagcca 60
ccatgccagt ctctaagtgc ccaaaaaagt cggagtccct gtggaagggg tgggaccgga 120
aggcccagag gaacggcctg cggagccagg tatacgctgt gaatggcgac tactatgtgg 180
gcgagtggaa ggacaacgtg aaacacggga aaggaacaca ggtctggaag aagaaaggag 240
ccatctatga gggggactgg aagtttggga agcgagacgg ctacggcacc ctcagccttc 300
ctgaccaaca gacaggaaag tgcaggagag tctactcagg ctggtggaaa ggtgataaga 360
aatcgggtta tgggatccag tttttcggac ccaaggagta ttatgagggt gactggtgtg 420
gcagccagcg cagcgggtgg ggccgcatgt attacagcaa cggcgacatc tacgagggac 480
agtgggagaa cgacaagccc aacggggagg gcatgctgcg cctgaagaac gggaaccgct 540
acgagggctg ctgggagaga ggcatgaaga acggggcggg gcgtttcttc catctggacc 600
acggccagct gtttgaaggc ttctgggtgg acaatatggc caaatgcggg acgatgatcg 660
actttggccg tgacgaggcc cctgagccca ctcagttccc cattcctgag gtcaaaatcc 720
tagaccctga tggtgtgctg gcggaggcct tggccatgtt caggaagaca gaggaaggag 780
attgatgcca gagaacacaa acgcttcagg agaaattcaa gcctgtgtca cccgatcgct 840
cagaccagtg cggctctggc tggaggagtc agcagcagct ccaggcatga ccccggcacc 900
ctcatagggc ccctcactac ccccagcact gggtcatttc ttgccaatag gaaggctggt 960
gcttctctcc caggctgtcc tcgggaccct cttcattctc tgatctcatc ctggaatgca 1020
tgagaataaa gaataaccaa gtggtaaaaa aaaa 1054

<210> 38
<211> 3360
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4841542CB1

<400> 38
```
agtgtgctgg aaaagcgctt cagccctccc cgcacagcct actgattccc ctgccgccct 60
tgctcacctc ctgctcgcca tggagtcgct ggttttcgcg cggcgctccg gccccactcc 120
ctcggcgcag agctagcccg gccgctggcg gaagggctga tcaagtcgcc caagcccccta 180
atgaagaagc aggcggtgaa gcggcaccac cacaagcaca acctgcggca ccgctacgag 240
ttcctggaga ccctgggcaa aggcacctac gggaaggtga agaaggcgcg ggagagctcg 300
gggcgcctgg tggccatcaa gtcaatccgg aaggacaaaa tcaaagatga gcaagatctg 360
atgcacatac ggagggagat tgagatcatg tcatcgctca accaccctca catcattgcc 420
atccatgaag tgtttgagaa cagcagcaag atcgtgatcg tcatggagta tgccagccgg 480
ggcgaccttt atgactacat cagcgagcgg cagcagctca gtgagcgcga agctaggcat 540
ttcttccggc agatcgtctc tgccgtgcac tattgccatc agaacagagt tgtccaccga 600
gatctcaagc tggagaacat cctcttgggt gccaatggga atatcaagat tgctgacttc 660
ggcctctcca acctctacca tcaaggcaag ttcctgcaga cattctgtgg gagccccctc 720
tatgcctcgc cagagattgt caatgggaag ccctacacag gcccagaggt ggacagctgg 780
tccctgggtg ttctcctcta catcctggtg catggcacca tgccctttga tgggcatgac 840
cataagatcc tagtgaaaca gatcagcaac ggggcctacc gggagccacc taaaccctct 900
gatgcctgtg gcctgatccg gtggctgttg atggtgaacc ccacccgccg ggccaccctg 960
gaggatgtgg ccagtcactg gtgggtcaac tggggctacg ccacccgagt gggagagcag 1020
gaggctccgc atgagggtgg gcaccctggc agtgactctg cccgcgcctc catggctgac 1080
tggctccggc gttcctcccg ccccctcctg gagaatgggg ccaaggtgtg cagcttcttc 1140
aagcagcatg cacctggtgg gggaagcacc acccctggcc tggagcgcca gcattcgctc 1200
aagaagtccc gcaaggagaa tgacatggcc cagtctctcc acagtgacac ggctgatgac 1260
actgcccatc gcgctggcaa gagcaacctc aagctgccaa agggcattct caagaagaag 1320
gtgtcagcct ctgcagaagg ggtacaggag gaccctccgg agctcagccc aatccctgcg 1380
agcccagggc aggctgcccc cctgctcccc aagaagggca ttctcaagaa gccccgacag 1440
cgcgagtctg gctactactc ctctcccgag cccagtgaat ctggggagct cttggacgca 1500
ggcgacgtgt ttgtgagtgg ggatcccaag gagcagaagc ctccgcaagc ttcagggctg 1560
ctcctccatc gcaaaggcat cctcaaactc aatggcaagt tctcccagac agccttggag 1620
ctcgcggccc ccaccacctt cggctccctg gatgaactcg ccccacctcg cccccctggcc 1680
cgggccagcc gaccctcagg ggctgtgagc gaggacagca tcctgtcctc tgagtccttt 1740
gaccagctgg acttgcctga acggctccca gagcccccac tgcggggctg tgtgtctgtg 1800
gacaacctca cggggcttga ggagcccccc tcagagggcc ctggaagctg cctgaggcgc 1860
tggcggcagg atcctttggg ggacagctgc ttttccctga cagactgcca ggaggtgaca 1920
gcgacctacc gacaggcact gagggtctgc tcaaagctca cctgagtgga gtaggcattg 1980
ccccagcccg gtcaggctct cagatgcagc tggttgcacc ccgaggggag atgccttctc 2040
ccccacctcc caggacctgc atcccagctc agaaggctga gagggtttgc agtggagccc 2100
tgagcagggc tggatatggg aagtaggcaa atgaaatgcg ccaagggttc agtgtctgtc 2160
ttcagccctg ctgaacgaag aggatactaa agagagggga acgggaatgc ccgcgacaga 2220
gtccacattg cctgtttctt gtgtacatgg gggggccaca gagacctgga aagagaactc 2280
tcccagggcc catctcctgc atcccatgaa tactctgtac acatggtgcc ttctaaggac 2340
agctccttcc ctactcattc cctgcccaag tggggccaga cctctttaca cacacattcc 2400
cgttcctacc aaccaccaga actggatggt ggcaccccta atgtgcatga ggcatcctgg 2460
gaatggtctg gagtaacgct tcgttatttt tatttttatt tttatttatt tatttatttt 2520
tttgagacgg agtttcgctc ttggtgccca ggctagagtg caatggcgcg atctcagctc 2580
acctcaacct ccgcctcccg ggttcaagcg attctcctgc ctcagcctcc ctagtagctg 2640
ggattacagg cgcccgccac catgcccggc taattttgta tttttagtag agacagggtt 2700
tctccatgtt ggtcaggctg gtctcaaact cccgacctca ggtgatccac ccacctcggc 2760
ctcccaaagt gctgggatta caggcgtgag ccaccgcgcc ccacctaacc cttccttatt 2820
tagcctagga gtaagagaac acaatctctg tttcttcaat ggttctcttc ccttttccat 2880
cctccaaacc tggcctgagc ctcctgaagt tgctgctgtg aatctgaaag acttgaaaag 2940
cctccgcctg ctgtgtggac ttcatctcaa ggggcccagc ctcctctgga ctccaccttg 3000
gacctcagtg actcagaact tctgcctcta agctgctcta aagtccagac tatggatgtg 3060
ttctctaggc cttcaggact ctagaatgtc catatttatt tttatgttct tggctttgtg 3120
ttttaggaaa agtgaatctt gctgttttca ataatgtgaa tgctatgttc tgggaaaatc 3180
cactatgaca tctaagtttt gtgtacagag agatattttt gcaactattt ccacctcctc 3240
ccacaacccc ccacactcca ctccacactc ttgagtctct ttacctaatg gtctctacct 3300
aatggacctc cgtggccaaa aagtaccatt aaaaccagaa aggtgattgg aaaaaaaaaa 3360
```

<210> 39
<211> 2240
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472695CB1

<400> 39
```
cgggctgaaa agtttctccc ggtgcagaat tccgggctca gcgacagcct gcgccgagtg 60
tgcgcacctg tcggagaccc gccagtccgc cggccccggc ctgaagttaa atcattttgg 120
aaagtgatac agcaaaacaa gggttcctcc agttttggtg gtggaaatgt cacagacatc 180
aagcattggt agtgcagaat ctttaatttc actggagaga aaaaaagaaa aaaatatcaa 240
cagagatata acctccagga aagatttgcc ctcaagaacc tcaaatgtag agagaaaagc 300
atctcagcaa caatggggtc ggggcaactt tacagaagga aaagttcctc acataaggat 360
tgagaatgga gctgctattg aggaaatcta tacctttgga agaatattgg gaaaagggag 420
ctttggaata gtcattgaag cgacagacaa ggaaacagaa acgaagtggg caattaaaaa 480
agtgaacaaa gaaaaggctg gaagctctgc tgtgaagtta cttgaacgag aggtgaacat 540
tctgaaaagt gtaaaacatg aacacatcat acatctggaa caagtatttg aaacgccaaa 600
gaaaatgtac cttgtgatgg agctttgtga ggatggagaa ctcaaagaaa ttctggatag 660
gaaagggcat ttctcagaga atgagacaag gtggatcatt caaagtctcg catcagctat 720
agcatatctt cacaataatg atattgtaca tagagatctg aaactggaaa atataatggt 780
taaaagcagt cttattgatg ataacaatga aataaactta aacataaagg tgactgattt 840
tggcttagcg gtgaagaagc aaagtaggag tgaagccatg ctgcaggcca catgtgggac 900
tcctatctat atggcccctg aagttatcag tgcccacgac tatagccagc agtgtgacat 960
ttggagcata ggcgtcgtaa tgtacatgtt attacgtgga gaaccaccct ttttggcaag 1020
ctcagaagag aagcttttg agttaataag aaaaggagaa ctacattttg aaaatgcagt 1080
ctggaattcc ataagtgact gtgctaaaag tgttttgaaa caacttatga aagtagatcc 1140
tgctcacaga atcacagcta aggaactact agataaccag tggttaacag gcaataaact 1200
ttcttcggtg agaccaacca atgtattaga gatgatgaag gaatggaaaa ataacccaga 1260
aagtgttgag gaaaacacaa cagaagagaa gaataagccg tccactgaag aaaagttgaa 1320
aagttaccaa ccctggggaa atgtccctga tgccaattac acttcagatg aagaggagga 1380
aaaacagtct actgcttatg aaaagcaatt tcctgcaacc agtaaggaca actttgatat 1440
gtgcagttca agtttcacat ctagcaaact ccttccagct gaaatcaagg gagaaatgga 1500
gaaaacccct gtgactccaa gccaaggaac agcaaccaag taccctgcta aatccggcgc 1560
cctgtccaga accaaaaaga aactctaagg ttccctccag tgttggacag tacaaaaaca 1620
aagctgctct tgttagcact ttgatgaggg ggtaggaggg gaagaagaca gccctatgct 1680
gagcttgtag ccttttagct ccacagagcc ccgccatgtg tttgcaccag cttaaaattg 1740
aagctgctta tctccaaagc agcataagct gcacgtggca ttaaaggaca gccaccagta 1800
ggcttggcag tgggctgcag tggaaatcaa ctcaagatgt acacgaaggt tttttagggg 1860
ggcagatacc ttcaatttaa ggctgtgggc acacttgctc atttttactt caaattctta 1920
tgtttaggca cagctattta taggggaaaa caagaggcca aatatagtaa tggaggtgcc 1980
aaataattat gtgcactttg cactagaaga ctttgttaga aaattactaa taaacttgcc 2040
atacgtatta cagcagaagt gcttcagtca ttcacatgtg ttcgtgagat tttaggttgc 2100
tatagattgt ttaagacagc ttattttaaa tgtagaaaaa taggagattt tgtaactgct 2160
tgccattaac ttgctgctaa attcccaatg tattgattaa atcaataaaa aacagatgtt 2220
actcagcaaa aaaaaaaaaa                                             2240
```

<210> 40
<211> 3340
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477966CB1

<400> 40
```
cactataacc tttctctagg gtcaaagaga tgatgagtga caccagcacg ttccccaatc 60
acccttcctc ccctgctgca tccccatctg ggggaagggg agtcatggcc agccctgctt 120
gggacaggag caaagggtgg tcccagaccc ctcagagagc tgactttgtc tctacccccct 180
tgcaggttca tactctcagg ccagagaacc tcctgctggt gtccaccttg gatggaagtc 240
tccacgcact aagcaagcag acaggggacc tgaagtggac tctgagggat gatcccgtca 300
tcgaaggacc aatgtacgtc acagaaatgg cctttctctc tgacccagca gatggcagcc 360
tgtacatctt ggggacccaa aaacaacagg gattaatgaa actgccattc accatccctg 420
agctggttca tgcctctccc tgccgcagct ctgatggggt cttctacaca ggccggaagc 480
```

aggatgcctg gtttgtggtg gaccctgagt caggggagac ccagatgaca ctgaccacag 540
agggtccctc caccccccgc ctctacattg gccgaacaca gtatacggtc accatgcatg 600
acccaagagc cccagccctg cgctggaaca ccacctaccg ccgctactca gcgcccccca 660
tggatggctc acctgggaaa tacatgagcc acctggcgtc ctgcgggatg ggcctgctgc 720
tcactgtgga cccaggaagc gggacggtgc tgtggacaca ggacctgggc gtgcctgtga 780
tgggcgtcta cacctggcac caggacggcc tgcgccagct gccgcatctc acgctggctc 840
gagacactct gcatttcctc gccctccgct ggggccacat ccgactgcct gcctcaggcc 900
cccgggacac agccaccctc ttctctacct tggacaccca gctgctaatg acgctgtatg 960
tggggaagga tgaaactggc ttctatgtct ctaaagcact ggtccacaca ggagtggccc 1020
tggtgcctcg tggactgacc ctggcccccg cagatggccc caccacagat gaggtgacac 1080
tccaagtctc aggagagcga gagggctcac ccagcactgc tgttagatac cctcaggca 1140
gtgtggccct cccaagccag tggctgctca ttggacacca cgagctaccc ccagtcctgc 1200
acaccaccat gctgagggtc catcccaccc tggggagtgg aactgcagag acaagacctc 1260
cagagaatac ccaggcccca gccttcttct tggagctatt gagcctgagc cgagagaaac 1320
tttgggactc cgagctgcat ccagaagaaa aaactccaga ctcttacttg ggctgggac 1380
cccaagacct gctggcagct agcctcactg ctgtcctcct gggagggtgg attctctttg 1440
tgatgaggca gcaacagccg caggtggtgg agaagcagca ggagaccccc ctggcacctg 1500
cagactttgc tcacatctcc caggatgccc agtccctgca ctcgggggcc agccggagga 1560
gccagaagag gcttcagagt ccctcaaagc aagcccagcc actcgacgac cctgaagctg 1620
agcaactcac cgtagtgggg aagatttcct tcaatcccaa ggacgtgctg ggccgcgggg 1680
caggcgggac tttcgttttc cggggacagt ttgagggacg ggcagtggct gtcaagcggc 1740
tcctccgcga gtgctttggc ctggttcggc gggaagttca actgctgcag gagtctgaca 1800
ggcaccccaa cgtgctccgc tacttctgca ccgagcgggg accccagttc cactacattg 1860
ccctggagct ctgccgggcc tccttgcagg agtacgtaga aaacccggac ctggatcgcg 1920
gggtctggga gcccgaggtc gtgctgcagc agctgatgtc tggcctggcc cacctgcact 1980
ctttacacat agtgcaccgg gacctgaagc caggaaatat tctcatcacc gggcctgaca 2040
gccagggcct gggcagagtg gtgctctcag acttcggcct ctgcaagaag ctgcctgctg 2100
gccgctgtag cttcagcctc cactccggca tccccggcac ggaaggctgg atggcgcccg 2160
agcttctgca gctcctgcca ccagacagtc ctaccagcgc tgtggacatc ttctctgcag 2220
gctgcgtgtt ctactacgtg ctttctggtg gcagccaccc ctttggagac agtctttatc 2280
gccaggcaaa catcctcaca ggggctccct gtctggctca cctggaggaa gaggtccacg 2340
acaaggtggt tgcccgggac ctggttggag ccatgttgag cccactgccg cagccacgcc 2400
cctctgcccc ccaggtgctg gcccacccct tcttttggag cagagccaag caactccagt 2460
tcttccagga cgtcagtgac tggctggaga aggagtccga gcaggagccc ctggtgaggg 2520
cactggaggc gggaggctgc gcagtggtcc gggacaactg gcacgagcac atctccatgc 2580
cgctgcagac agatctgaga aagttccggt cctataaggg gacatcagtg cgagacctgc 2640
tccgtgctgt gaggaacaag aagcaccact acagggagct cccagttgag gtgcgacagg 2700
cactcggcca agtccctgat ggcttcgtcc agtacttcac aaaccgcttc ccacggctgc 2760
tcctccacac gcaccgagcc atgaggagct gcgcctctga gagcctcttc ctgccctact 2820
acccgccaga ctcagaggcc aggaggccat gccctggggc cacagggagg tgaggtgggc 2880
tggatgccac acagatggtc tccgtgctgg ctcactgaag agctgagcct gtggctggcc 2940
tcagaatcag gctgggtgca gtggctcaca cctgtaatcc cagcattttg ggaggctgag 3000
tgagaggatc acttgagctc aggagttcga gaccagcctg gccaacatgg caacacccca 3060
tttctacaaa aaatttgtaa aattagccag gcatggtggc gcacgcctgt agtcccagct 3120
gcttgggagg ctgaggtggg agaatcactt gagcccagga gttcgaggct gcagtgagcc 3180
aggatcatgc cactgcactc cagcctggtc cacagagaga cactgtcacc ccctttcccc 3240
cacaagactg gcagaggctg ggcagcctgg ggctgatgaa gcagagatgt tcgctggatc 3300
ccagctcctg gcacactgta aggaaataca acgaagaggt 3340

<210> 41
<211> 2539
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7163416CB1

<400> 41
cggaggactg gcccagcaag gtcccaggtc ttccctctcc tcagcgccta agagagaggc 60
ccagtgcggg tgaggagtcg cgaggaagag gcggaaggcg ccggaaggca ccatgttccg 120
caagaaaaag aagaaacgcc ctgagatctc agcgccacag aacttccagc accgtgtcca 180
cacctccttc gaccccaaag aaggcaagtt tgtgggcctc cccccacaat ggcagaacat 240
cctggacaca ctgcggcgcc ccaagcccgt ggtggaccct tcgcgaatca cacgggtgca 300
gctccagccc atgaagacag tggtgcgggg cagcgcgatg cctgtggatg gctacatctc 360
ggggctgctc aacgacatcc agaagttgtc agtcatcagc tccaacaccc tgcgtggccg 420
cagccccacc agccggcggc gggcacagtc cctggggctg ctgggggatg agcactgggc 480

caccgaccca gacatgtacc tccagagccc ccagtctgag cgcactgacc cccacggcct 540
ctacctcagc tgcaacgggg gcacaccagc aggccacaag cagatgccgt ggcccgagcc 600
acagagccca cgggtcctgc ccaatgggct ggctgcaaag gcacagtccc tgggccccgc 660
cgagtttcag ggtgcctcgc agcgctgtct gcagctgggt gcctgcctgc agagctcccc 720
accaggagcc tcgcccccca cgggcaccaa taggcatgga atgaaggctg ccaagcatgg 780
ctctgaggag gcccggccac agtcctgcct ggtgggctca gccacaggca ggccaggtgg 840
ggaaggcagc cctagcccta agacccggga gagcagcctg aagcgcaggc tattccgaag 900
catgttcctg tccactgctg ccacagcccc tccaagcagc agcaagccag gccctccacc 960
acagagcaag cccaactcct ctttccgacc gccgcagaaa gacaaccccc caagcctggt 1020
ggccaaggcc cagtccttgc cctcggacca gccggtgggg accttcagcc ctctgaccac 1080
ttcggatacc agcagccccc agaagtccct ccgcacagcc ccggccacag gccagcttcc 1140
aggccggtct tccccagcgg gatccccccg cacctggcac gcccagatca gcaccagcaa 1200
cctgtacctg ccccaggacc ccacggttgc caagggtgcc ctggctggtg aggacacagg 1260
tgttgtgaca catgagcagt tcaaggctgc gctcaggatg gtggtggacc agggtgaccc 1320
ccggctgctg ctggacagct acgtgaagat tggcgagggc tccaccggca tcgtctgctt 1380
ggcccgggag aagcactcgg gccgccaggt ggccgtcaag atgatggacc tcaggaagca 1440
gcagcgcagg gagctgctct tcaacgaggt ggtgatcatg cgggactacc agcacttcaa 1500
cgtggtggag atgtacaaga gctacctggt gggcgaggag ctgtgggtgc tcatggagtt 1560
cctgcaggga ggagccctca cagacatcgt ctcccaagtc aggctgaatg aggagcagat 1620
tgccactgtg tgtgaggctg tgctgcaggc cctggcctac ctgcatgctc agggtgtcat 1680
ccaccgggac atcaagagtg actccatcct gctgaccctc gatggcaggg tgaagctctc 1740
ggacttcgga ttctgtgctc agatcagcaa agacgtccct aagaggaagt ccctggtggg 1800
aacccctac tggatggctc ctgaagtgat ctccaggtct ttgtatgcca ctgaggtgga 1860
tatctggtct ctgggcatca tggtgattga gatggtagat ggggagccac cgtacttcag 1920
tgactcccca gtgcaagcca tgaagaggct ccgggacagc cccccaccca agctgaaaaa 1980
ctctcacaag gtcagttggc acacaagggt gcgacctcgc agaccccatt cctcctgagg 2040
caaggggacc agaacctggg ctcccagcat ctcccttcca ctgaagccac agggtctggg 2100
ctcctggaaa aggctcctct ttccccacac aaaacccgca cctgggtgtg gagccgcatc 2160
tacgcacaag ttcgcatgtg cgctccgaca agtcgcctcc cacggctgtg gcaggagagt 2220
tgctgcttgg cagaagggtt gctgcttggc aggcactggt cggaagccca gtggggccca 2280
tgagcaggga aagccaggac accagcaatc cctgctgtcc agggagggat ccggagaagc 2340
ttcactgagc acaaaccctt ctaacccgtg tcgggagatc cataccatga ttcgatgtcc 2400
tgtccatcac ggcgagtcgg ctcatgctcc atcgttgcac accccgacac agctaagcca 2460
cagcgttccc cttaaagcca gtataagtgc atggaagtgt atacatgtaa cccttttgc 2520
caaatcggcc ccaacccg 2539

<210> 42
<211> 2377
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472822CB1

<400> 42
agtgtgctgg aaagttgaat tggaattccc tgtggctgtc cgaaggcagg gtgtccggag 60
agcggtgggc tgacctgttc ctacaccttg catcatgcca gctttgtcaa cgggatctgg 120
gagtgacact ggtctgtatg agctgttggc tgctctgcca gcccagctgc agccacatgt 180
ggatagccag gaagacctga ccttcctctg ggatatgttt ggtgaaaaaa gcctgcattc 240
attggtaaag attcatgaaa aactacacta ctatgagaag cagagtccgg tgcccattct 300
ccatggtgcg gcggccttgg ccgatgatct ggccgaagag cttcagaaca agccattaaa 360
cagtgagatc agagagctgt tgaaactact gtcaaaaccc aatgtgaagg ctttgctctc 420
tgtacatgat actgtggctc agaagaatta cgacccagtg ttgcctccta tgcctgaaga 480
tattgacgat gaggaagact cagtaaaaat aatccgtctg gtcaaaaata gagaaccact 540
gggagctacc attaagaagg atgaacagac cggggcgatc attgtggcca gaatcatgag 600
aggaggagct gcagatagaa gtggtcttat tcatgttggt gatgaactta gggaagtcaa 660
cgggatacca gtggaggata aaaggcctga ggaaataata cagattttgg ctcagtctca 720
gggagcaatt acatttaaga ttatacccgg cagcaaagag gagacaccat caaaagaagg 780
caagatgttt atcaaagccc tctttgacta taatcctaat gaggataagg caattccatg 840
taaggaagct gggctttctt tcaaaaaggg agatattctt cagattatga gccaagatga 900
tgcaacttgg tggcaagcga aacacgaagc tgatgccaac cccagggcag gcttgatccc 960
ctcaaagcat ttccaggaaa ggagattggc tttgagacga ccagaaatat tggttcagcc 1020
cctgaaagtt tccaacagga aatcatctgg ttttagaaaa agttttcgtc ttagtagaaa 1080
agataagaaa acaaataaat ccatgtatga atgcaagaag agtgatcagt acgacacagc 1140
tgacgtaccc acatacgaag aagtgacacc gtatcggcga caaactaatg aaaaatacag 1200
actcgttgtc ttggttggtc ccgtgggagt agggctgaat gaactgaaac gaaagctgct 1260

gatcagtgac acccagcact atggcgtgac agtgccccat accaccagag caagaagaag 1320
ccaggagagt gatggtgttg aatacatttt catttccaag catttgtttg agacagatgt 1380
acaaaataac aagtttattg aatatggaga atataaaaac aactactacg gcacaagtat 1440
agactcagtt cggtctgtcc ttgctaaaaa caaagtttgt ttgttggatg ttcagcctca 1500
tacagtgaag catttaagga cactagaatt taagccctat gtgatattta taaagcctcc 1560
atcaatagag cgtttgagag aaacaagaaa aaatgcaaag attatttcaa gcagagatga 1620
ccaaggtgct gcaaaaccct tcacagaaga agattttcaa gaaatgatta aatctgcaca 1680
gataatggaa agtcaatatg gtcatctttt tgacaaaatt ataataaatg atgacctcac 1740
tgtggcattc aatgagctca aaacaacttt tgacaaatta gagacagaga cccattgggt 1800
gccagtgagc tggttacatt cataactaag agaaatttcc ataattgtct ttttctatag 1860
agtgcatgat gaaatcaatt acagttttgg tagtagggtt tttaaatcta tatcactgtc 1920
atagatgtac aatcttggtt caagttgaat gctggttttg tttgtatctt tttacagcct 1980
tatttcaaac gccatgtgtt agtataagat ccgaaatcaa aatatgcaca gtactgtatt 2040
ctaagcaaaa cctcaaacct tctcgttgtc ttcaatatcg ctctatctcc aagatgaggc 2100
tgaaattttc agagagactt agctagaggc ttagtatgta tgggagttca gcgcttctgc 2160
tggtctcagg tgtggctgct gctgtcgagt ttgcatgtta gctgttgaag gtatcaattc 2220
agcagccatg agcagctcca gacagacagc gtgagctctg ctgtttctgg gtggatcatc 2280
acagatttag ccgggcaggc agtaaggtgt cctcttacta ttcaaaagtg tagactttct 2340
gggggatcca ctagttctac acgccgcccc cgtgacc 2377

<210> 43
<211> 2897
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477486CB1

<400> 43
atggtggcgg ggttaacttt ggggaagggc ccggagtccc cggatggtga tgtcagcgtg 60
ccggagagaa aggacgaggt ggcgggggga ggcggagagg aggaggaggc cgaagagaga 120
gggcgccacg cccaatatgt gggcccctat cggctggaga agacgctggg caaaggacag 180
acagggctgg ttaaactcgg ggtccactgc atcacgggtc agaaggtcgc catcaagatc 240
gtgaaccggg agaagctgtc ggagtcggtg ctgatgaagg tggagcggga gatcgccatc 300
ctgaagctca tcgaacaccc acatgtcctc aagctccacg acgtctacga gaacaagaaa 360
tatttgtacc tggttctgga gcacgtctcg gggggtgagc tattcgacta cctggtaaag 420
aaggggagac tgacgcccaa ggaggcccga aagttcttcc gccagattgt gtctgcgctg 480
gacttctgcc acagctactc catctgccac agagacctaa agcccgagaa cctgcttttg 540
gatgagaaaa acaacatccg cattgcagac ttcggcatgg cgtccctgca ggtgggggac 600
agcctcctgg agaccagctg cgggtccccc cattatgcgt gtccagaggt gattaagggg 660
gaaaaatatg atggccgccg ggcagacatg tggagctgtg gagtcatcct cttcgccctg 720
ctcgtggggg ctctgccctt tgatgacgac aacctccgcc agctgctgga gaaggtgaaa 780
cggggcgtct tccacatgcc ccacttcatt cctccagatt gccagagcct cctgagggga 840
atgatcgaag tggagcccga aaaaaggctc agtctggagc aaattcagaa acatccttgg 900
tacctaggcg ggaaacacga gccagacccg tgcctggagc cagcccctgg ccgccgggta 960
gccatgcgga gcctgccatc caacggagag ctggaccccg acgtcctaga gagcatggca 1020
tcactgggct gcttcaggga ccgcgagagg ctgcatcgcg agctgcgcag tgaggaggag 1080
aaccaagaaa agatgatata ttatctgctt ttggatcgga aggagcggta tcccagctgt 1140
gaggaccagg acctgcctcc ccggaatgat gttgaccccc cccggaagcg tgtggattct 1200
cccatgctga gccgtcacgg gaagcggcga ccagagcgga agtccatgga agtcctgagc 1260
atcaccgatg ccgggggtgg tggctcccct gtacccaccc gacgggcctt ggagatggcc 1320
cagcacagcc agagatcccg tagcgtcagt ggagcctcca cgggtctgtc ctccagccct 1380
ctaagcagcc caaggagtcc ggtcttttcc ttttcaccgg agccgggggc tggagatgag 1440
gctcgaggcg gggggctccc gacttccaaa acgcagacgc tgccttctcg ggggccccagg 1500
ggtgggggcg ccggggagca gcccccgccc cccagtgccc gctccacacc cctgcccggc 1560
cccccaggct ccccgcgctc ctctggcggg accccccttgc actcgcctct gcacacgccc 1620
cgggccagtc ccaccgggac cccggggaca acaccacccc ccagccccgg cggtggcgtc 1680
gggggagccg cctggaggag tcgtctcaac tccatccgca acagcttcct gggctcccct 1740
cgctttcacc ggcgcaagat gcaggtccct accgctgagg agatgtccag cttgacgcca 1800
gagtcctccc cggagctggc aaaacgctcc tggttcggga acttcatctc cttggacaaa 1860
gaagaacaaa tattcctcgt gctaaaggac aaacctctca gcagcatcaa agcagacatc 1920
gtccatgcct ttctgtcgat ccccagcctg agtcacagtg tgctgtcaca gaccagcttc 1980
agggccgagt acaaggccag tggcggcccc tccgtcttcc aaaagcccgt ccgcttccag 2040
gtggacatca gctcctctga gggtccagag ccctccccgc gacgggacgg cagcggaggt 2100
ggtggcatct actccgtcac cttcactctc atctcgggtc ccagccgtcg gttcaagcga 2160
gtggtggaga ccatccaggc acagctcctg agcactcatg accagccctc cgtgcaggcc 2220

```
ctggcagacg agaagaacgg ggcccagacc cggcctgctg gtgccccacc ccgaagcctg 2280
cagcccccac ccggccgccc agacccagag ctgagcagct ctccccgccg aggccccccc 2340
aaggacaaga agctcctggc caccaacggg acccctctgc cctgacccca cggggccggg 2400
gagggagggg accccccctcc acccccttc cgtgcccccc aactgtgaat ctgtaaataa 2460
ggcccaagga acatgtcggg agggggtgg acacaaaaac cggccttgcc ctgcagggat 2520
ggggctccac aggccgtgcc caactgcggg tggttctagg ggaacagggg gcggggggagc 2580
tgtttctatt ttatttattg attaatttat tattttattt attgatcaat ctctctgcgg 2640
ggtgcggtgg gggagggacg ggagctggtt ggggtggctt agcagatccg gacagggccc 2700
tctgtccctg tgtcgtcccc aacccctct tcccgggccc ctcctcccct ggtccttccc 2760
cccacgacct tctgtacgga tttgctctcc ggaaggaatt ctgataacgc gtgatcctgc 2820
ctgcgtccgt gtctctgatt ccgccggcgg caaaaaaaac acaacaccaa caacacaaca 2880
gggcacaaca aaaaaaa                                                 2897
```

<210> 44
<211> 3361
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3773709CB1

<220>
<221> unsure
<222> 96
<223> a, t, c, g, or other

<400> 44

```
ggctgagccg ggttggggcc cgggttgggc cgcccgggga ctctggagca ttgggatttg 60
tagcgcgccc tctgggtagg cggctgtagc ggagangcgt gcgggatcgg gatgtcgggg 120
ctgctcacgg acccggagca gagagcgcag gagccgcggt accccggctt cgtgctgggg 180
ctggatgtgg gcagttctgt gatccgctgc cacgtctatg accgggcggc gcgggtctgc 240
ggctccagcg tgcagaaggt agaaaatctt tatcctcaaa ttggctgggt agaaattgat 300
cctgatgttc tttggattca atttgttgcc gtaataaaag aagcagtcaa agctgcagga 360
atacagatga atcaaattgt tggtcttggc atttcaacac agagagcaac ttttattacg 420
tggaacaaga aaacaggaaa tcattttcac aactttataa gttggcaaga cttaagagct 480
gttgaacttg taaaatcttg gaataattct cttcttatga agatatttca cagttcttgc 540
cgagtgcttc actttttcac tagaagtaaa cgacttttta cagccagttt gttcactttc 600
acaacccagc agacttcttt gagattggtc tggattttac agaacttgac tgaggtgcaa 660
aaggcagttg aagaagaaaa ttgctgcttt gggactatcg atacctggtg gttatataag 720
ctcacaaaag gttctgtata tgccacagat ttttcaaatg ctagtacaac tggacttttt 780
gacccatata gccacaattt tggatcagtg gatgaagaga tatttggtgt gcctatacca 840
atagttgcct tggttgctga ccagcaatca gccatgtttg gagagtgctg cttccagaca 900
ggtgatgtga aattaaccat gggaactggg acatttttgg atattaacac tggaaatagc 960
cttcaacaga ctactggagg cttttatcca ttaattgggt ggaagattgg gcaagaagtc 1020
gtatgcttag ctgaaagcaa tgcaggagac actggtactg ccataaaatg ggctcagcag 1080
ttagaccttt tcacagatgc tgctgagact gaaaaaatgg ccaaaagttt ggaggattct 1140
gaaggagttt gttttgttcc atcttttagt ggattacagg ctccattaaa tgacccctgg 1200
gcatgtgcct cttttatggg tttgaagcct tctaccagta aataccatct tgtacgagca 1260
atattggagt caatagcttt cagaaacaaa cagttatatg agatgatgaa gaaagagatt 1320
catattcctg taagaaaaat ccgggcagat ggaggagttt gtaagaatgg ttttgtcatg 1380
cagatgactt cagacctgat taatgagaat atagacagac ctgccgacat tgacatgtca 1440
tgcctgggtg cagcttctct agctggcctt gctgttgggt tttggactga caaggaggaa 1500
ctaaagaaac tgagacaaag tgaagtggtt ttcaagccac agaagaaatg tcaagaatat 1560
gaaatgagtc tggaaaactg ggccaaagca gtgaaacgct ccatgaattg gtataacaag 1620
acataacact aaatgaaatg atcaaaacca taggtagctg gtttatgtga cgtgcagatg 1680
agatgaagct cagggataac ccatatgaca atgactaaga ggagaaaatt ttaaataagc 1740
ttcataactt aagaagcatt gcttttaaaa aaacaaaacg gaacaaaaaa ctcttatttt 1800
tttcccctaa accatggtaa ggcagcaata cctcaaaact ttatatcttc tattttgtag 1860
caaattccaa aggacattag tcatttccaa ccacattttg acagttatgg gtcctcttcc 1920
tttttatact gggtcagtgg tacataggaa cataatgatt taccatccaa gctaatagtt 1980
ctgggtcaag taccatgcac atattgttcc aaaattatgt gaaacgtatt tctttaattc 2040
tttaagtggg ctatttgaag tacatatagc taaaaagaaa gaataactga gaaaatgtgg 2100
aattttgaaa cattaatatt ttatgtttaa agccataatt tcctaatatt atatccaaat 2160
atgagcttaa tatgtccctc tcagataagc ttatgagata gttaatgctt tcctttactg 2220
gtcttaaaga cactgcctta atttttcctt gttcaaccaa aatctgagca ttctttctat 2280
gttgaaaaca ctgaaaaact aattttagtt aatgaactag aaagaatatt gttttttaag 2340
```

aaacagaaaa atactactta ttttccttct caaataacgt ttctttcaaa aacttctggc 2400
tgaagtataa catgctggta gttaacataa atcttgtctt tctcttgttc tttatctttc 2460
tttgttattt agatgcttgt ataaatgtct tttgttttta ttaagtgcct aattgacaga 2520
gcttaatttg aagaagtgcc ctaatttatt gaccacttaa gaattgcctt tattggggta 2580
ttttatttgt tcctgcgtct ttttgatgtt tgttcagtct actcatccct gtgagtatgt 2640
gtgggggaca gctgatagaa gggaggagag tgtgtctatg ctcaggattg ccctttagcc 2700
actcagccag agatccacag ggagcaacaa ggacagtttc acatgcttag actttcttgg 2760
aagaaacagt gaggaggagt aagtcgtgag tagtgtcaag ctggatgtag aattgtccta 2820
aggcagttga ccccaccttc caacatgttt tcactttatt tgcccctccc tacatttggg 2880
ttaggttcca tttggatttg cagcaataat gactttattt ctctcttggt caggatttgg 2940
cacataaaat ccttttatta tagaactagc tattttagtt acatagtaat gtaactaatg 3000
gagagattta tagagaattt tgtttttgct gtcatatatg tccattttgg agacagatat 3060
gatagaacta gaaattaagt tgcatttctg caagtgccat ttgaatgaac ttcaagtatc 3120
ttcttaatta ttaaattttc tgatgaaggc attgtaacaa atatatagta ttattaaatc 3180
taattaatat ttggaaatat taataaatag gtattttatt tactgtaaaa agtcaaactt 3240
cattatgtag ataaatctta ttcttttcat tctttcccct gtttacatcc tttttacaaa 3300
gcttagtcac caattaaagc tttcctatca aaatcagaaa agaaaaaaag agaagacaca 3360
c 3361

<210> 45
<211> 1662
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7477204CB1

<400> 45
atggtggaca tgggggccct ggacaacctg atcgccaaca ccgcctacct gcaggcccgg 60
aagccctcgg actgcgacag caaagagctg cagcggcggc ggcgtagcct ggccctgccc 120
gggctgcagg gctgcgcgga gctccgccag aagctgtccc tgaacttcca cagcctgtgt 180
gagcagcagc ccatcggtcg ccgcctcttc cgtgacttcc tagccacagt gcccacgttc 240
cgcaaggcgg caaccttcct agaggacgtg cagaactggg agctggccga ggagggaccc 300
accaaagaca gcgcgctgca ggggctggtg gccacttgtg cgagtgcccc tgccccgggg 360
aacccgcaac ccttcctcag ccaggccgtg gccaccaagt gccaagcagc caccactgag 420
gaagagcgag tggctgcagt gacgctggcc aaggctgagg ccatggcttt cttgcaagag 480
cagcccttta aggatttcgt gaccagcgcc ttctacgaca agtttctgca gtggaaactc 540
ttcgagatgc aaccagtgtc agacaagtac ttcactgagt tcagagtgct ggggaaaggt 600
ggttttgggg aggtatgtgc cgtccaggtg aaaaacactg ggaagatgta tgcctgtaag 660
aaactggaca agaagcggct gaagaagaaa ggtggcgaga agatggctct cttggaaaag 720
gaaatcttgg agaaggtcag cagccctttc attgtctctc tggcctatgc ctttgagagc 780
aagacccatc tctgccttgt catgagcctg atgaatgggg gagacctcaa gttccacatc 840
tacaacgtgg gcacgcgtgg cctggacatg agccgggtga tcttttactc ggcccagata 900
gcctgtggga tgctgcacct ccatgaactc ggcatcgtct atcgggacat gaagcctgag 960
aatgtgcttc tggatgacct cggcaactgc aggttatctg acctggggct ggccgtggag 1020
atgaagggtg gcaagcccat cacccagagg gctggaacca atggttacat ggctcctgag 1080
atcctaatgg aaaaggtaag ttattcctat cctgtggact ggtttgccat gggatgcagc 1140
atttatgaaa tggttgctgg acgaacacca ttcaaagatt acaaggaaaa ggtcagtaaa 1200
gaggatctga agcaaagaac tctgcaagac gaggtcaaat tccagcatga taacttcaca 1260
gaggaagcaa aagatatttg caggctcttc ttggctaaga aaccagagca acgcttagga 1320
agcagagaaa agtctgatga tcccaggaaa catcatttct ttaaaacgat caactttcct 1380
cgcctggaag ctggcctaat tgaacccccca tttgtgccag acccttcagt ggtttatgcc 1440
aaagacatcg ctgaaattga tgatttctct gaggttcggg gggtggaatt tgatgacaaa 1500
gataagcagt tcttcaaaaa ctttgcgaca ggtgctgttc ctatagcatg gcaggaagaa 1560
attatagaaa cgggactgtt tgaggaactg aatgacccca acagacctac gggttgtgag 1620
gagggtaatt catccaagtc tggcgtgtgt ttgttattgt aa 1662

<210> 46
<211> 3225
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3016969CB1

<400> 46
agtgtgctgg aaaggccgcc agggaggagc aggccaccct cctggccaaa gccccctcat 60
tcgagactgc cctccggctg cctgcctctg gcacccactt ggcccctggc cacagccact 120
ccctggaaca tgactctccg agcacccccc gcccctcctc ggaggcctgc ggtgaggcac 180
agcgactgcc ttcagccccc tccggggggg cccctatcag ggacatgggg caccctcagg 240
gctccaagca gcttccatcc actggtggcc acccaggcac tgctcagcca gagaggccat 300
ccccggacag cccttggggg cagccagccc ctttctgcca ccccaagcag ggttctgccc 360
cccaggaggg ctgcagcccc cacccagcag ttgccccatg ccctcctggc tccttccctc 420
caggatcttg caaagaggcc cccttagtac cctcaagccc cttcttgggg acagccccag 480
gcacccccctg cccctgccaa agcaagcccc ccattggact ctaagatggg gcctggagac 540
atctctcttc ctgggaggcc aaaacccggc ccctgcagtt ccccagggtc agcctcccag 600
gcgagctctt cccaagtgag ctccctcagg gtgggctcct cccaggtggg cacagagcct 660
ggcccctccc tggatgcgga gggctggacc caggaggctg aggatctgtc cgactccaca 720
cccaccttgc agcggcctca ggaacaggtg accatgcgca agttctccct gggtggtcgc 780
gggggctacg caggcgtggc tggctatggc acctttgcct ttggtggaga tgcaggggc 840
atgctggggc aggggcccat gtgggccagg atagcctggg ctgtgtccca gtcggaggag 900
gaggagcagg aggaggccag ggctgagtcc cagtcggagg agcagcagga ggccagggct 960
gagagcccac tgccccaggt cagtgcaagg cctgtgcctg aggtcggcag ggctcccacc 1020
aggagctctc cagagcccac cccatgggag gacatcgggc aggtctccct ggtgcagatc 1080
cgggacctgt caggtgatgc ggaggcggcc gacacaatat ccctggacat ttccgaggtg 1140
gaccccgcct acctcaacct ctcagacctg tacgatatca agtacctccc attcgagttt 1200
atgatcttca ggaaagtccc caagtccgct cagccagagc cgccctcccc catggctgag 1260
gaggagctgg ccgagttccc ggagcccacg tggccctggc caggtgaact gggcccccac 1320
gcaggcctgg agatcacaga ggagtcagag gatgtggacg cgctgctggc agaggctgcc 1380
gtgggcagga agcgcaagtg gtcctcgccg tcacgcagcc tcttccactt ccctgggagg 1440
cacctgccgc tggacgagcc tgcagagctg gggctgcgtg agagagtgaa ggcctccgtg 1500
gagcacatct cccggatcct gaagggcagg ccggaaggtc tggagaagga ggggcccccc 1560
aggaagaagc caggccttgc ttccttccgg ctctcaggtc tgaagagctg ggaccgagcg 1620
ccgacattcc taagggagct ctcagatgag actgtggtcc tgggccagtc agtgacactg 1680
gcctgccagg tgtcagccca gccagctgcc caggccacct ggagcaaaga cggagccccc 1740
ctggagagca gcagccgtgt cctcatctct gccaccctca agaacttcca gcttctgacc 1800
atcctggtgg tggtggctga ggacctgggt gtgtacacct gcagcgtgag caatgcgctg 1860
gggacagtga ccaccacggg cgtcctccgg aaggcagagc gcccctcatc ttcgccatgc 1920
ccggatatcg gggaggtgta cgcggatggg gtgctgctgg tctggaagcc cgtggaatcc 1980
tacggccctg tgacctacat tgtgcagtgc agcctagaag gcggcagctg gaccacactg 2040
gcctccgaca tctttgactg ctgctacctg accagcaagc tctcccgggg tggcacctac 2100
accttccgca cggcatgtgt cagcaaggca ggaatgggtc cctacagcag cccctcggag 2160
caagtcctcc tgggagggcc cagccacctg gcctctgagg aggagagcca ggggcggtca 2220
gcccaacccc tgcccagcac aaagaccttc gcattccaga cacagatcca gaggggccgc 2280
ttcagcgtgg tgcggcaatg ctgggagaag gccagcgggc gggcgctggc cgccaagatc 2340
atcccctacc accccaagga caagacagca gtgctgcgcg aatacgaggc cctcaagggc 2400
ctgcgccacc cgcacctggc ccagctgcac gcagcctacc tcagcccccg gcacctggtg 2460
ctcatcttgg agctgtgctc tgggcccgag ctgctcccct gcctggccga gagggcctcc 2520
tactcagaat ccgaggtgaa ggactacctg tggcagatgt tgagtgccac ccagtacctg 2580
cacaaccagc acatcctgca cctggacctg aggtccgaga acatgatcat caccgaatac 2640
aacctgctca aggtcgtgga cctgggcaat gcacagagcc tcagccagga gaaggtgctg 2700
ccctcagaca agttcaagga ctacctagag accatggctc cagagctcct ggagggccag 2760
ggggctgttc cacagacaga catctgggcc atcggtgtga cagccttcat catgctgagc 2820
gccgagtacc cggtgagcag cgagggtgca cgcgacctgc agagaggact gcgcaagggg 2880
ctggtccggc tgagccgctg ctacgcgggg ctgtccgggg gcgccgtggc cttcctgcgc 2940
agcactctgt gcgcccagcc ctggggccgg ccctgcgcgt ccagctgcct gcagtgcccg 3000
tggctaacag aggagggccc ggcctgttcg cggcccgcgc ccgtgacctt ccctaccgcg 3060
cggctgcgcg tcttcgtgcg caatcgcgag aagagacgcg cgctgctgta caagaggcac 3120
aacctggccc aggtgcgctg agggtcgccc cggccacacc cttggtctcc ccgctggggg 3180
tcgctgcaga cgcgccaata aaaacgcaca gccgggcgag aaaaa 3225

<210> 47
<211> 4772
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 063497CB1

<400> 47
gcggacggac gctcgcctgc cggctgagga aaaagaagca actaacaaaa cactgtgata 60

```
ataaggatta ttcagtatgc agtttgcagg atatccatga cgacattgaa aatgaatttt 120
ttgtattcac cagatattct tatatgagaa gatctatttt aaacagtcta aatatttttt 180
cttctgttgg accagcatgg caggatttaa gcgagggtat gatggaaaga ttgctggatt 240
atatgatctg gataaaacct tgggtcgagg ccattttgcc gtggttaaac ttgccaggca 300
tgtctttacg ggtgaaaagg tggcagtaaa agttattgac aagacaaaac tggacactct 360
agctactggt catcttttcc aggaagtgag atgcatgaaa ctagtgcagc atcctaacat 420
cgtccgcctt tatgaagtta ttgacaccca gaccaaacta tatcttattc tagaacttgg 480
ggatggagga gatatgtttg attatataat gaaacatgag gagggtctta atgaagactt 540
ggccaagaag tattttgctc agatagttca tgctatatct tattgccata aactccatgt 600
ggttcacaga gacttaaaac cagagaatgt agtcttcttt gaaaaacaag gtcttgtaaa 660
gttgacagac tttgggttca gcaacaaatt tcaaccaggg aagaagctca ctacaagctg 720
tggatctctt gcatattccg ctccagaaat tctgcttggt gatgagtatg atgcacctgc 780
agtagatatt tggagtctgg gagtgatcct tttcatgttg gtgtgtgggc agccgccctt 840
tcaagaagcc aatgacagtg aaacactgac aatgatcatg gattgcaaat atacagtacc 900
atcccatgtg tctaaagagt gtaaagacct aatcacacgg atgctacaga gagatcccaa 960
gagaagggct tctttagaag agattgaaaa tcatccttgg cttcagggag tggacccttc 1020
accagctaca aagtataaca ttccccttgt gtcatacaaa aatctctcgg aagaggagca 1080
caacagcatc attcagcgca tggtgcttgg ggacatagcg gatcgagacg ccattgtaga 1140
agccctggaa accaacaggt ataaccatat cacagccaca tacttcctgc tggctgaaag 1200
gatcctgaga gaaaagcaag agaaagaaat acagaccaga tctgcaagcc cgagcaatat 1260
caaggcccag tttaggcagt catggccaac caaaattgat gtaccccagg accttgagga 1320
tgacctcacg gccactcctt tgtcccacgc gactgtccct cagtctcctg ctcgggctgc 1380
tgacagtgtc ctcaatggcc acaggagcaa aggcctgtgt gactcagcta agaaagatga 1440
cctccctgag ttggctggac cagcactctc tacggtgcca cccgcaagct taaaacccac 1500
agccagtggg cggaagtgtc tgttcagggt ggaagaagat gaagaggaag atgaggagga 1560
caagaaaccc atgtccctct caacacaagt ggttttgcgc cggaagccat ctgtaaccaa 1620
ccgcctgaca tccaggaaga gtgcgcccgt cctcaaccag atctttgagg aaggggaatc 1680
tgacgatgag tttgacatgg atgagaatct gcctcccaag ttgagcaggt taaagatgaa 1740
tatagcttct ccaggtacag ttcacaaacg ctaccaccgg aggaaaagtc agggccgggg 1800
ctccagctgc agtagttcgg agaccagtga tgatgattct gaaagccggc ggcggctcga 1860
taaagatagc gggttcacct actcctggca ccgacgggat agcagcgagg ggcccccctgg 1920
cagtgagggg gatggcgggg gccagagcaa gccaagcaat gccagtggag gggtggacaa 1980
ggccagcccc agtgagaaca atgctggtgg gggcagtccc tccagcggct cgggtggcaa 2040
ccccaccaat acatcgggta ccacacgccg ctgtgccggc cccagcaact ccatgcagct 2100
ggcctctcgc agtgctgggg agctcgttga gagcctcaaa ctcatgagcc tctgcctcgg 2160
ctcccagctt catgggagca ccaagtacat tattgatcca cagaatggct tgtcattttc 2220
cagtgtgaaa gtccaagaga aatctacgtg gaaaatgtgc attagctcca cagggaatgc 2280
agggcaggtc cctgcagtgg gcggcataaa gtttttctct gaccacatgg cagataccac 2340
cactgaattg gaacggataa agagcaagaa cctgaaaaat aacgtgctgc agctacctct 2400
gtgcgaaaag accatctctg tgaacatcca gcggaaccct aaggaggggc tgctgtgcgc 2460
atccagccca gccagctgtt gccatgtcat ctgactgtgg ccccatctgg ccgctagcac 2520
gcttcctgct cagagcagtg aagaccggct cacttcactg ttccatttgg ttttactatt 2580
ttaaagtggg cgttaggagc aattatttat tacctttcca tttgttcgcc tgatgatgtg 2640
acaatgcatg gtctttgtgc atgctgctag acacttttct ttcccagccg aaaagcctat 2700
tatgtaattt ttacattcat aattttaatg tggatgatca ggattaaatc aagatatata 2760
tctggaacct cttaaaaatg gagcacttag aaatttgttg ttctgcactt aacctagaga 2820
gagaaaaaat gcttttcttt gtgaaaaatc tgaattcctg tcctgacctt ctgtgatgtg 2880
gaaaccctag gctctgagac acactctctg gtgtctgaga cagaaccaaa gcaataacgt 2940
tgtgatgccc acaggcctgg agccagctag cgaccttgtg ccgcccagct gtccatggcc 3000
cgtgcagagc agaggacagt gagtgtctgc actgagaacc ttaaaccaca gttgaacata 3060
cccacacctg tttgtcttaa gctatagtgt aaaaacaaag tttgggctct gaaaatttaa 3120
ctgaaaaaga tttccttgtt tttgtaatag gtgagataaa gtacttagat ttataaggca 3180
gcttcccctg tagtgataaa ttacaagcag acaatcttat tttgtaatgt gatgaagtga 3240
tgatgtctta actctactta gagagtgtat gtctgtctaa cagaacaaaa agatgctctg 3300
tgtaaattcc ttcctgtagg gcacactgca ggatttccat gtagatagaa gaactatagg 3360
gcctagtaca gaaggtgcac acaaatgttg gcaaagtcaa aaccccatga attaaaacct 3420
actggaattt ggtttttagg agtttggtaa ttagattatc tcttttgtta ttttcattca 3480
gttatatcct ttggctcagc tagctttgaa attggctgat gaaaaaatat acataaaagg 3540
gtaaaattca cacatacagc aaacaaaaat gcacaaagcc tgcttcgtaa cttttttttc 3600
tggaattgtt tttcactttg cctttttctg ccaaaacaat aatcaaagaa ctcttgcttt 3660
aacctattcc tgtacaaaga ctgtttttga ccagataatc atctgttgtg gcattctatc 3720
ttgtaggaca ctgtatattg caaattgctg attatggaag gggccagttg ctgtttttttc 3780
atgcagtgcc ctgggagtct taaaagcagt gcttagcaac attggtgata gcatgtggct 3840
gggacccagg gcccttcccc actcttcagc cccgagtcat gtgtctgagg tgacggactg 3900
agacgcatct ggtcctgtaa ttcagagagt gggcacatca ccaaagaact gcattgctgt 3960
ggtcactgtt tcttcaagta cacactgact ctgctacttt aggataaata tattttactc 4020
agaactctga atttcacagt atacttacta aactaagtaa aaatgatact taaaatactt 4080
```

attttacttt ctagacctag gctagatgtt ttaagctaca gctctagttc attgtgatat 4140
ttataatttg aaagctatga gaatagatgt gtgggtgaag ccatagaaca tatttgcttg 4200
aaattcttga gcagggatct tataaagggc cagaaataag atgtgtggtt cacatagata 4260
gtgagcgtaa catctgtatt aaacatagga gagaagttta taaagggcat tggcaataaa 4320
ctctttgttg cagctgtttt ccaagcagtg taaatacttt ttcctgtgat tatgtatagc 4380
cttggaatgg caccttttaa ctaacccata tgtgtttggt ttcaatggtt ttttatattc 4440
agatgtatat atggtgctca ctttaggatc agcagtgttg accatttatg ctgcatagct 4500
gtattatagc cttattagtt gtgtggttga cccttggggt atacaaaaat ctctcggaag 4560
aggagcacaa cagcatcatt cagcgcatgg tgcttgggga catagcggat cgagacgcca 4620
ttgtagaagc cctggaaacc aacaggtata accatatcac agccacatac ttcctgctgg 4680
ctgcaaagga tcctgagaga aaagcaagag aaagaaatac agaccagatc tgcaagcccg 4740
agcaatatca aggcccagtt taggcagtca tg 4772

<210> 48
<211> 1880
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1625436CB1

<400> 48
ctcttgctcc ctcggccggg cggcggtgac tgtgcaccga cgtcggcgcg ggctgcaccg 60
ccgcgtccgc ccgcccgcca gcatggccac caccgccacc tgcacccgtt tcaccgacga 120
ctaccagctc ttcgaggagc ttggcaaggg tgctttctct gtggtccgca ggtgtgtgaa 180
gaaaacctcc acgcaggagt acgcagcaaa aatcatcaat accaagaaat tgtctgcccg 240
ggatcaccag aaactagaac gtgaggctcg gatatgtcga cttctgaaac atccaaacat 300
cgtgcgcctc catgacagta tttctgaaga agggtttcac tacctcgtgt ttgaccttgt 360
taccggcggg gagctgtttg aagacattgt ggccagagag tactacagtg aagcagatgc 420
cagccactgt atacatcaga ttctggagag tgttaaccac atccaccagc atgacatcgt 480
ccacagggac ctgaagcctg agaacctgct gctggcgagt aaatgcaagg gtgccgccgt 540
caagctggct gattttggcc tagccatcga agtacaggga gagcagcagg cttggtttgg 600
ttttgctggc accccaggtt acttgtcccc tgaggtcttg aggaaagatc cctatggaaa 660
acctgtggat atctgggcct gcggggtcat cctgtatatc ctcctggtgg gctatcctcc 720
cttctgggat gaggatcagc acaagctgta tcagcagatc aaggctggag cctatgattt 780
cccatcacca gaatgggaca cggtaactcc tgaagccaag aacttgatca accagatgct 840
gaccataaac ccagcaaagc gcatcacggc tgaccaggct ctcaagtacc cgtgggtctg 900
tcaacgatcc acggtggcat ccatgatgca tcgtcaggag actgtggagt gtttgcgcaa 960
gttcaatgcc cggagaaaac tgaagggtgc catcctcacg accatgcttg tctccaggaa 1020
cttctcagtt ggcaggcaga gctccgcccc cgcctcgcct gccgcgagcg ccgccggcct 1080
ggccgggcaa gctgccaaaa gcctattgaa caagaagtcg gatggcggtg tcaagaaaag 1140
gaagtcgagt tccagcgtgc acctaatgcc acagagcaac aacaaaaaca gtctcgtaag 1200
cccagcccaa gagcccgcgc ccttgcagac ggccatggag ccacaaacca ctgtggtaca 1260
caacgctaca gatgggatca agggctccac agagagctgc aacaccacca cagaagatga 1320
ggacctcaaa gctgccccgc tccgcactgg gaatggcagc tcggtgcctg aaggacggag 1380
ctcccgggac agaacagccc cctctgcagg catgcagccc cagccttctc tctgctcctc 1440
agccatgcga aaacaggaga tcattaagat tacagaacag ctgattgaag ccatcaacaa 1500
tggggacttt gaggcctaca cgaagatttg tgatccaggc ctcacttcct ttgagcctga 1560
ggcccttggt aacctcgtgg aggggatgga tttccataag ttttactttg agaatctcct 1620
gtccaagaac agcaagccta tccataccac catcctaaac ccacacgtcc acgtgattgg 1680
ggaggacgca gcgtgcatcg cctacatccg cctcacccag tacatcgacg ggcagggtcg 1740
gcctcgcacc agccagtcag aagagacccg ggtctggcac cgtcgggatg gcaagtggct 1800
caatgtccac tatcactgct caggggcccc tgccgcaccg ctgcagtgag ctcagccaca 1860
ggggctttag gagattccag 1880

<210> 49
<211> 5747
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3330646CB1

<400> 49
ggtaggcagg cggctgagcc ggcggcgggt ggcctgccca acgtgtgctg ggtgggagaa 60

ggcgaggcgg cagcgatgct gtctcttccg tgaggagcgc agaggaggtc gcggcgccgg 120
aggccccaga aggctcgaag gcgccgcggg ctggggtcgg tggcttaggg agcccgtccg 180
gccatggtgg ccgcgggtgg tggttggcgc ggctgcgctg cggcccgggg cagtgcggag 240
ccgggacagt cgcggcgctg acgcccgcgg gccccagctg cagatatgaa gcggagccgc 300
tgccgcgacc gaccgcagcc gccgccgccc gaccgccggg aggatggagt tcagcgggca 360
gcggagctgt ctcagtcttt gccgccgcgc cggcgagcgc cgcccgggag gcagcggctg 420
gaggagcgga cgggccccgc ggggcccgag ggcaaggagc aggatgtagc aactggagtt 480
agtcccctgc tcttcaggaa actcagtaat cctgacatat tttcatccac tggaaaagtt 540
aaacttcagc gacaactgag tcaggatgat tgtaagttat ggagaggaaa cctggccagc 600
tctctatcgg gtaagcagct gctccctttg tccagcagtg tacatagcag tgtgggacag 660
gtgacttggc agtcgtcagg agaagcatca aacctggttc gaatgagaaa ccagtccctt 720
ggacagtctg caccttctct tactgctggc ctgaaggagt tgagccttcc aagaagaggc 780
agcttttgtc ggacaagtaa ccgcaagagc ttgattgtga cctctagcac atcacctaca 840
ctaccacggc cacactcacc actccatggc cacacaggta acagtccttt ggacagcccc 900
cggaatttct ctccaaatgc acctgctcac ttttcttttg ttcctgcccg taggactgat 960
gggcggcgct ggtctttggc ctctttgccc tcttcaggat atggaactaa cactcctagc 1020
tccactgtct catcatcatg ctcctcacag gaaaagctgc atcagttgcc tttccagcct 1080
acagctgatg agctgcactt tttgacgaag catttcagca cagagagcgt accagatgag 1140
gaaggacggc agtccccagc catgcggcct cgctcccgga gcctcagtcc cggacgatcc 1200
ccagtatcct ttgacagtga aataataatg atgaatcatg tttacaaaga aagattccca 1260
aaggccaccg cacaaatgga agagcgacta gcagagttta tttcctccaa cactccagac 1320
agcgtgctgc ccttggcaga tggagccctg agctttattc atcatcaggt gattgagatg 1380
gcccgagact gcctggataa atctcggagt ggcctcatta catcacaata cttctacgaa 1440
cttcaagaga atttggagaa acttttacaa gatgctcatg agcgctcaga gagctcagaa 1500
gtggcttttg tgatgcagct ggtgaaaaag ctgatgatta tcattgcccg cccagcacgt 1560
ctcctggaat gcctggagtt tgaccctgaa gagttctacc acctttaga agcagctgag 1620
ggccacgcca aagagggaca agggattaaa tgtgacattc cccgctacat cgttagccag 1680
ctgggcctca cccgggatcc cctagaagaa atggcccagt tgagcagctg tgacagtcct 1740
gacactccag agacagatga ttctattgag ggccatgggg catctctgcc atctaaaaag 1800
acaccctctg aagaggactt cgagaccatt aagctcatca gcaatggcgc ctatggggct 1860
gtatttctgg tgcggcacaa gtccacccgg cagcgctttg ccatgaagaa gatcaacaag 1920
cagaacctga tcctacggaa ccagatccag caggccttcg tggagcgtga catactgact 1980
ttcgctgaga accccttgt ggtcagcatg ttctgctcct ttgataccaa gcgccacttg 2040
tgcatggtga tggagtacgt tgaaggggga gactgtgcca ctctgctgaa gaatattggg 2100
gccctgcctg tggacatggt gcgtctatac tttgcggaaa ctgtgctggc cctggagtac 2160
ttacacaact atggcatcgt gcaccgtgac ctcaagcctg acaacctcct aattacatcc 2220
atggggcaca tcaagctcac ggactttgga ctgtccaaaa tgggcctcat gagtctgaca 2280
acgaacttgt atgagggtca tattgaaaag gatgcccggg aattcctgga caagcaggta 2340
tgcgggaccc cagaatacat tgcgcctgag gtgatcctgc gccagggcta tgggaagcca 2400
gtggactggt gggccatggg cattatcctg tatgagttcc tggtgggctg cgtccctttt 2460
tttggagata ctccggagga gctctttggg caggtgatca gtgatgagat tgtgtggcct 2520
gagggtgatg aggcactgcc cccagacgcc caggacctca cctccaaact gctccaccag 2580
aaccctctgg agagacttgg cacaggcagt gcctatgagg tgaagcagca cccattcttt 2640
actggtctgg actggacagg acttctccgc cagaaggctg aatttattcc tcagttggag 2700
tcagaggatg atactagcta ttttgacacc cgctcagagc gataccacca catggactcg 2760
gaggatgagg aagaagtgag tgaggatggc tgccttgaga tccgccagtt ctcttcctgc 2820
tctccaaggt tcaacaaggt gtacagcagc atggagcggc tctcactgct cgaggagcgc 2880
cggacaccac ccccgaccaa gcgcagcctg agtgaggaga aggaggacca ttcagatggc 2940
ctggcagggc tcaaaggccg agaccggagc tgggtgattg gctcccctga gatattacgg 3000
aagcggctgt cggtgtctga gtcatcccac acagagagtg actcaagccc tccaatgaca 3060
gtgcgacgcc gctgctcagg cctcctggat gcgcctcggt tcccggaggg ccctgaggag 3120
gccagcagca ccctcaggag gcaaccacag gagggtatat gggtcctgac acccccatct 3180
ggagaggggg tatctggccc tgtcactgaa cactcagggg agcagcggcc aaagctggat 3240
gaggaagctg ttggccggag cagtggttcc agtccagcta tggagacccg aggccgtggg 3300
acctcacagc tggctgaggg agccacagcc aaggccatca gtgacctggc tgtgcgtagg 3360
gcccgccacc ggctgctctc tggggactca acagagaagc gcactgctcg ccctgtcaac 3420
aaagtgatca agtccgcctc agccacagcc ctctcactcc tcattccttc ggaacaccac 3480
acctgctccc cgttggccag ccccatgtcc ccacattctc agtcgtccaa cccatcatcc 3540
cgggactctt ctccaagcag ggacttcttg ccagcccttg gcagcatgag gcctcccatc 3600
atcatccacc gagctggcaa gaagtatggc ttcaccctgc gggccattcg cgtctacatg 3660
ggtgactccg atgtctacac cgtgcaccat atggtgtggc acgtggagga tggaggtccg 3720
gccagtgagg cagggcttcg tcaaggtgac ctcatcaccc atgtcaatgg ggaacctgtg 3780
catggcctgg tgcacacgga ggtggtagag ctgatcctga agagtggaaa caaggtggcc 3840
atttcaacaa ctcccctgga gaacacatcc attaaagtgg ggccagctcg gaagggcagc 3900
tacaaggcca agatggcccg aaggagcaag aggagccgcg gcaaggatgg gcaagaaagc 3960
agaaaaagga gctccctgtt ccgcaagatc accaagcaag catccctgct ccacaccagc 4020
cgcagccttt cttcccttaa ccgctccttg tcatcagggg agagtggcc aggctctccc 4080

acacacagcc acagcctttc cccccgatct cccactcaag gctaccgggt gacccccgat 4140
gctgtgcatt cagtgggagg gaattcatca cagagcagct cccccagctc cagcgtgccc 4200
agttccccag ccggctctgg gcacacacgg cccagctccc tccacggtct ggcacccaag 4260
ctccaacgcc agtaccgctc tccacggcgc aagtcagcag gcagcatccc actgtcacca 4320
ctggcccaca ccccttctcc cccaccccca acagcttcac ctcagcggtc cccatcgccc 4380
ctgtctggcc atgtagccca ggcctttccc acaaagcttc acttgtcacc tcccctgggc 4440
aggcaactct cacggcccaa gagtgcggag ccaccccgtt caccactact caagagggtg 4500
cagtcggctg agaaactggc agcagcactt gccgcctctg agaagaagct agccacttct 4560
cgcaagcaca gccttgacct gccccactct gaactaaaga aggaactgcc gcccagggaa 4620
gtgagccctc tggaggtagt tggagccagg agtgtgctgt ctggcaaggg ggccctgcca 4680
gggaaggggg tgctgcagcc tgctccctca cgggccctag gcaccctccg gcaggaccga 4740
gccgaacgac gggagtcgct gcagaagcaa gaagccattc gtgaggtgga ctcctcagag 4800
gacgacaccg aggaagggcc tgagaacagc cagggtgcac aggagctgag cttggcacct 4860
cacccagaag tgagccagag tgtggcccct aaaggagcag gagagagtgg ggaagaggat 4920
cctttcccgt ccagagaccc taggagcctg ggcccaatgg tcccaagcct attgacaggg 4980
atcacactgg ggcctcccag aatggaaagt cccagtggtc cccacaggag gctcgggagc 5040
ccacaagcca ttgaggaggc tgccagctcc tcctcagcag gccccaacct aggtcagtct 5100
ggagccacag accccatccc tcctgaaggt tgctggaagg cccagcacct ccacacccag 5160
gcactaacag cactttctcc cagcacttcg ggactcaccc ccaccagcag ttgctctcct 5220
cccagctcca cctctgggaa gctgagcatg tggtcctgga aatcccttat tgagggccca 5280
gacagggcat ccccaagcag aaaggcaacc atggcaggtg ggctagccaa cctccaggat 5340
ttggaaaaca caactccagc ccagcctaag aacctgtctc ccagggagca ggggaagaca 5400
cagccaccta gtgcccccag actggcccat ccatcttatg aggatcccag ccagggctgg 5460
ctatgggagt ctgagtgtgc acaagcagtg aaagaggatc cagccctgag catcacccaa 5520
gtgcctgatg cctcaggtga cagaaggcag gacgttccat gccgaggctg cccccctcacc 5580
cagaagtctg agcccagcct caggaggggc caagaaccag gggccatca aaagcatcgg 5640
gatttggcat tggttccaga tgagctttta aagcaaacat agcagttgtt tgccatttct 5700
tgcactcaga cctgtgtaat atatgctcct ggaaaccaaa aaaaaaa 5747

<210> 50
<211> 3418
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3562763CB1

<400> 50
gaggtgggac gccccgcggc ctacgctcct ggcctccccg ccttggcctg gccgtttaac 60
cgattctttc gcccgcaggt cacaatccaa ggtccggctc ctccgcgtcc cagggccgga 120
cggagggatg aggcaggggg ggcccgggca gcgccgttgc tgctcccccc gccgcccgca 180
gccatggaaa cggggaagga cggcgcccgc agaggtacac aaagcccgga gcggaaaatg 240
cgaagcccag tgccgcgggc gcccagcacg aagctgagcc ggcggcggcg cccgggccat 300
ggatccggtg gctgccgagg ccccgggcga ggccttcctg gcgcggcgac ggcctgaggg 360
cggtggcggg tccgcgcggc cgcgttacag cctgttggcg gagatcgggc gcggcagcta 420
cggcgtggtt tatgaggcag tggccgggcg cagcggggcc cgggtggcgg tcaagaagat 480
ccgctgcgac gcccccgaga acgtggagct ggcgctggct gaattctggg ccctcaccag 540
cctcaagcgg cgccaccaga acgtcgtgca gtttgaggag tgcgtcctgc agcgcaatgg 600
gttagcccag cgcatgagtc acggcaacaa gagctcgcag ctttacctgc gcctggtgga 660
gacctcgctg aaaggagaaa ggatcctggg ttatgctgag gagccctgct atctctggtt 720
tgtcatggag ttctgtgaag gtggagacct gaatcagtat gtcctgtccc ggaggccaga 780
cccagccacc aacaaaagtt tcatgctaca gctgacgagc gccattgcct tcctgcacaa 840
aaaccatatt gtgcacaggg acctgaagcc agacaacatc ctcatcacag agcggtctgg 900
cacccccatc ctcaaagtgg ccgactttgg actaagcaag gtctgtgctg ggctggcacc 960
ccgaggcaaa gagggcaatc aagacaacaa aaatgtgaat gtgaataagt actggctgtc 1020
ctcagcctgc ggttcggact tctacatggc tcctgaagtc tgggagggac actacacagc 1080
caaggcggac atctttgccc tgggcattat catctgggca atgatagaaa gaatcacttt 1140
tattgactct gagaccaaga aggagctcct ggggacctac attaaacagg ggactgagat 1200
cgtccctgtt ggtgaggcgc tgctagaaaa cccaaagatg gagttgcaca tcccccaaaa 1260
acgcaggact tccatgtctg aggggatcaa gcagctcttg aaagatatgt tagctgctaa 1320
cccacaggac cggcctgatg cctttgaact tgaaaccaga atggaccagg tcacatgtgc 1380
tgcttaaaat tcagggctaa gcattttggg tgattttaaa ctaggtcgat tcctcgggac 1440
ccacagtctc accacgtctc ctccagagga cggcagaggg tacaggtggt ggcctggccg 1500
gttggcgatc tcccgacagc tggatccggc aatgtgaagc ttttgtttgg gtttccccgc 1560
ttctttttag ttttgcttta ttttttttcct tttcttttct tttttttttt tcctctttcc 1620
ttttttttaaa tttaaaccat tgagacttca gaagagcagg acacaatgct gtggacaggc 1680

accaatttct ttaaagaaat tcaatgtggg caaggcatat gtgtaaattt cacttttact 1740
ttttataagg ggttagggag ctatttttgg ttttgtcctt cactttccct ctgtcttcct 1800
tctttatact tttctcagtt ctacttatga cacctcattt ccctagagaa ggcctgcctc 1860
cccataggga atctgggggt ttcttctgga acggggcgtg aggacacaag gaggcctctg 1920
ggccacgcct ccctaccaga tgcaggaact cctggactcc ttggtgggct ggccctggct 1980
agcccttggg cctcggagat gatcagaggt gaagaaccgc ctggaagagg acaggcccag 2040
ggtttggcca ggagaactaa gaaggtctca actccaggct ttgttgtgtt taagctattg 2100
agagccccag gccacaccag gacttgcagt ggtgggaatc cattcctctt ctgccctgtg 2160
ttgcagggaa ctaggaggta agggtggagg gcgaccatct cgctcttgct ggcggtggag 2220
cagccatccc tgcctttctg ttgggaaaaa ctgttgtgcc aaactcttgt gtggaacaca 2280
gctgggtctt cagcaggcat ctgtcactgc cgtgaggtca gcgcttctca cctaactgcc 2340
tcctggattg tcatcttccc agatgtgtcc catagtgtcc aggtgtcaca gagacggcct 2400
gaggccctaa gatctggttg tgactttgcc atgataacag ggtgtcctga actggctgcc 2460
gttgtcgtgt tctcacagtg aagggcgtgc cctgtgtgcc ggggtccatg gtgtcatatg 2520
cagtgacaca cactgtcaag cgccatttcc ctcacccctg gagacttact gttaggtgcc 2580
tgccctcagt atagacgtat ccaatgggaa aacagcggac ctgcccagag cagggaggtg 2640
tcgtggaact gggtagaccc cctgcagcgt taggggccca tttgtgggct cgccaccttc 2700
aggcttcccc agccatgaca cttcagcccc gccacccatg cctgtctgct gcagccatcc 2760
ttgcactctc cagcgacact ctcgcacctc cctaggggaa gcttccctcc ccctgggctg 2820
ctgctctgag cccgtctgtt ctccccctgc aagaaggggc aatgctcttg tgttgtccct 2880
ctgtctggac gcgcctggcc actccgaagg cttttcaccc cattatggcc aaatagtata 2940
gggccactgg ggagggggaa gggaatcatt ttgtgttcat ttttgttttc tgtttcacct 3000
aaaccagcat aggattgata ggggagacgg ttggcgggca tttccgtttc tatgtgacta 3060
tgtgaccaag gcagcagggg cttttacctg ctaggcggca gtcctttggc cctgagaatt 3120
tgggagagaa cagtgcatca ggccaggctc agcaatatgt ttgctcacat tctttcagcc 3180
ttctctcacc cccctcaaca ccaaactttc ttccttgtga gcagaaggtt ggctgctgtt 3240
agcaggatcc cacagtgata accaggccct tcccttccta agccaaaacc cattgtgact 3300
gcctgtctct cctgtctctg acttctcagg cagcctcctg agtgcactga gttgtatccg 3360
agagggtggg aacagcagca tcccctaatt gcagtacacg gttccttttc cgcccgcc 3418

<210> 51
<211> 995
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 621293CB1

<400> 51
cactttgact ggccacccga atctgaaatc cagaaccgtc tcatggtgcc agaggacatc 60
tcagagctgg agacggctca gaaactgctg gagtatcata ggaacatcgt cagggtcatt 120
ccctcctacc ccaaaatcct caaagtcatc agtgctgacc agccatgtgt ggacgtcttc 180
taccaggctc tgacctatgt ccaaagcaac catcgtacta atgccccgtt cacccccgagg 240
gtgctgctgc tcgggcctgt gggcagtggg aaaagtctgc aggccgccct cctggcccag 300
aaatacaggc ttgtcaatgt ctgctgtggg caactgctga aagaggctgt ggcagatagg 360
accacgtttg gcgagctcat ccagcccttc tttgaaaagg agatggcagt tcctgacagc 420
ctcctcatga aggtgctgag ccagcgcctg gaccagcagg actgcatcca gaaaggctgg 480
gtgctacacg gcgtcccgcg ggacctcgac caggcacacc tgctgaaccg cctgggctac 540
aatcccaaca gggtgttttt cctgaatgtg ccatttgatt ccatcatgga gcggctgact 600
ctgagaagaa ttgatccagt cactggggaa aggtaccacc tcatgtacaa gccacctccc 660
accatggaga tccaggctcg cctcctgcag aacccaaagg atgctgaaga gcaggtcaag 720
ctgaaaatgg acctgttcta caggaactca gctgacttgg agcagttgta tgggtcggcc 780
atcaccctca atggggacca ggacccatac acagtcttcg aatacatcga gagtgggatc 840
attaatcccc tgcccaagaa aatcccctga tgggttcaga gccaggagcg ctgccccagg 900
gaaagagtta atcccctgcc cccagccccc cagcctcggc acagctcccc taaaaagcca 960
ataaagcctg ctggatacca aaaaaaaaaa aaagc 995

<210> 52
<211> 2459
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480774CB1

<400> 52
```
gcggcgtagt ggttctgaac atggatagga ggggagatga tagctgctgg cgtccggtga 60
gcgtgggcag agcgtagtgc gggcagctgc ccagcggaag gatcggatga gactggaggc 120
gccgcgagga gggcggcggc ggcagccggg acagcagcga cctgggcccg gcgcaggggc 180
cccggcgggg cggccggagg gggggggcc ctgggcccgg acagaggagt ccagcctcca 240
cagcgagcct gagagggccg gcctcgggcc tgcgccgggg acagagagtc cgcaggcaga 300
attctggaca gacggacaga ctgagcccgc ggcagctggc cttggagtag agaccgagag 360
gcccaagcaa aagacggagc cagacaggtc cagcctccgg acgcatctag aatggagctg 420
gtcagagctg gagacgactt gtctttggac ggagaccggg acagatggcc tttggactga 480
tccgcacagg tccgacctcc agtttcagcc cgaggaggcc agcccctgga cacagccagg 540
ggttcatggg ccctggacag agctggaaac gcatgggtca cagactcagc cagagagggt 600
caagtcctgg gctgataacc tctggaccca ccagaacagt tccagcctcc agactcaccc 660
agaaggagcc tgtccctcaa aagagccaag tgctgatggc tcctggaaag aattgtatac 720
tgatggctcc aggacacaac aggatattga aggtccctgg acagagccat atactgatgg 780
ctcccagaaa aaacaggata ctgaagcagc caggaaacag cctggcactg gtggtttcca 840
aatacaacag gatactgatg gctcctggac acaacctagc actgacggtt cccagacagc 900
acctgggaca gactgcctct tgggagagcc tgaggatggc ccattagagg aaccagagcc 960
tggagaattg ctgactcacc tgtactctca cctgaagtgt agccccctgt gccctgtgcc 1020
ccgcctcatc attacccctg agaccccctga gcctgaggcc cagccagtgg gaccccccctc 1080
ccgggttgag gggggcagcg gcggcttctc ctctgcctct tctttcgacg agtctgagga 1140
tgacgtggtg gccgggggcg gaggtgccag cgatcccgag gacaggtctg ggagcaaacc 1200
ctggaagaag ctgaagacag ttctgaagta ttcacccttt gtggtctcct tccgaaaaca 1260
ctacccttgg gtccagcttt ctggacatgc tgggaacttc caggcaggag aggatggtcg 1320
gattctgaaa cgtttctgtc agtgtgagca gcgcagcctg gagcagctga tgaaagaccc 1380
gctgcgacct ttcgtgcctg cctactatgg catggtgctg caggatggcc agaccttcaa 1440
ccagatggaa gacctcctgg ctgactttga gggcccctcc attatggact gcaagatggg 1500
cagcaggacc tatctggaag aggagctagt gaaggcacgg gaacgtcccc gtccccggaa 1560
ggacatgtat gagaagatgg tggctgtgga ccctggggcc cctacccctg aggagcatgc 1620
ccagggtgca gtcaccaagc cccgctacat gcagtggagg gaaaccatga gctccacctc 1680
taccctgggc ttccggatcg agggcatcaa gaaggcagat gggacctgta acaccaactt 1740
caagaagacg caggcactgg agcaggtgac aaaagtgctg gaggacttcg tggatggaga 1800
ccacgtcatc ctgcaaaagt acgtggcatg cctagaagaa cttcgtgaag ctctggagat 1860
ctccccttc ttcaagaccc acgaggtggt aggcagctcc ctcctcttcg tgcacgacca 1920
caccggcctg gccaaggtct ggatgataga cttcggcaag acggtggcct tgcccgacca 1980
ccagacgctc agccacaggc tgccctgggc tgagggcaac cgtgaggacg gctacctctg 2040
gggcctggac aacatgatct gcctcctgca ggggctggca cagagctgag ctgctcagcc 2100
accatcaggt taattggatg gcgccagtct ggctggagga gccctgagat gccatgggag 2160
gcctgaggtt ggccacgggg gagctggcct ccagggacgg gagagattgt gtcatgtgcc 2220
acacgagacc aacgtggaaa agtctgaagg gccttgggag accaggtagc acctggcccc 2280
atcatgatgc aggggttttg gggacctgga aggaaggtga tgaggcagtg agtcagaaaa 2340
accagaacgg ggtccccgga tctgccggga aggcttctga ggggctgccc gtgagagcat 2400
tcagttcaca tgtaacaggg tagggggatc cactagttta taatgccggc cgcgtggta 2459
```